PRR

UNIVERSITY OF MICHIGAN

1955

READING ROOM

1831

SED · OMNIA · DISPOSVISTI

YORK · OXFORD · CAMBRIDGE · EDINBURGH · DUBLIN · BRISTOL · LIVERPOOL · NEWCASTLE · BIRMINGHAM · GLASGOW · PLYMOUTH · MANCHESTER · CORK · SOUTHAMPTON · SWANSEA · IPSWICH · BELFAST · HULL · CHELTENHAM · LEEDS · BATH · NOTTINGHAM · DUNDEE · NORWICH · EXETER · BRIGHTON · BRADFORD · SHEFFIELD · SOUTHPORT · MONTREAL · CARDIFF · TORONTO · DOVER · S. AFRICA · WINNIPEG · LEICESTER · PORTSMOUTH · AUSTRALIA · BOURNEMOUTH · LONDON · ABERDEEN · BLACKPOOL

The ADVANCEMENT OF SCIENCE

Vol. XII, No. 45

Techniques in Archaeology
Animal Behaviour
Water Control for Increased Crop Production
Modern Theories of Cosmology
Friction
Localisation in the Cerebral Cortex
Electricity Production in Great Britain
Milk Liberation
The Association of Psychology with other Disciplines
Heat in the Earth
John Phillips
Biology of Sewers
Obesity
Megalosaurus

Published by the

BRITISH ASSOCIATION
for the Advancement of Science
Burlington House, London, W.1.

June 1955 Tel.: Regent 2109

Price: Seven Shillings and Sixpence

British Mosses and Liverworts

E. VERNON WATSON

A sound introduction to the bryophytes for the beginner and a useful companion for the experienced student. 154 mosses and 55 liverworts are described and drawn, with keys and notes on many others. There is a detachable field key. *45s. net*

Bird Navigation

G. V. T. MATTHEWS

A Manx Shearwater, carried from its nest in Skokholm to Boston U.S.A., and then released, was back at its nest twelve and a half days later. How did it find its way? Dr Matthews brings several years' research to bear in the problems of bird navigation in this Cambridge Monograph in Experimental Biology. *12s. 6d. net*

CAMBRIDGE UNIVERSITY PRESS

ADLARD & SON LTD.

BARTHOLOMEW PRESS, DORKING, SURREY

Telephone : Dorking 3874

PRINTERS TO SCIENTIFIC SOCIETIES

Experienced in mathematical formulae composition

CIMBRA

(Comércio Império Britânico-Brasil)

IMPORTERS OF

ROCK CRYSTAL QUARTZ

FOR PIEZO ELECTRIC AND OPTICAL PURPOSES

We are at your service PLEASE CONSULT US

CIMBRA LIMITED

123 PALL MALL, LONDON, S.W.1 WHITEHALL 9661 CIMBRALON, PICCY, LONDON

15 m.
m. I
Q
41
.B86
A31
v. 12

1831

1955

THE ADVANCEMENT OF SCIENCE

VOL. XII, No. 45 JUNE, 1955

CONTENTS

	PAGE
Officers and Council	2
Techniques in Archaeology, *by* dr. k. p. oakley, dr. e. t. hall, and dr. g. w. dimbleby	3
Animal Behaviour, *by* dr. n. tinbergen, dr. r. a. hinde and prof. a. rex knight	17
The Division for the Social and International Relations of Science	28
Water Control for Increased Crop Production, *by* dr. h. l. penman	29
Modern Theories of Cosmology, *by* mr. h. bondi and dr. d. w. sciama	33
The Nature of Friction Between Solids, *by* dr. f. p. bowden, f.r.s.	43
Localisation in the Cerebral Cortex : Report of a Symposium	49
The Association of Psychology with other Disciplines, *by* prof. f. v. smith	68
The Salient Geographical Features of Electricity Production in Great Britain, *by* e. m. rawstron	73
Milk Liberation, *by* prof. h. d. kay, c.b.e., f.r.s.	83
Heat in the Earth, *by* d. w. allan	89
John Phillips and the Early Meetings of the British Association, *by* j. m. edmonds and p. a. beardmore	97
The Biology of Sewers, *by* prof. e. a. spaul	105
Obesity : Report of a Symposium	115
Research Committees	128
Megalosaurus, the Oxford Dinosaur, *by* dr. w. e. swinton	130

Q
41
.B86
A31
V.12

OFFICERS AND COUNCIL OF THE BRITISH ASSOCIATION, 1955

PATRON.

HER MAJESTY THE QUEEN.

PRESIDENT, 1955.

Prof. Sir ROBERT ROBINSON, O.M., F.R.S.

GENERAL TREASURER.

M. G. BENNETT.

GENERAL SECRETARIES.

Sir RICHARD SOUTHWELL, F.R.S. | Dr. GEORGE TAYLOR.

ORDINARY MEMBERS OF THE COUNCIL.

R. J. BARTLETT.
Prof. L. F. BATES, F.R.S.
M. C. BURKITT.
P. RITCHIE CALDER, C.B.E.
Prof. H. G. CHAMPION, C.I.E.
Prof. G. R. CLEMO, F.R.S.
Prof. H. P. GILDING.
Dr. D. B. HARDEN.
A. C. HARTLEY, C.B.E.
Prof. H. R. HEWER.
Prof. K. S. ISLES.
Dr. WILLIS JACKSON, F.R.S.
A. GRAY JONES.
Sir HAROLD SPENCER JONES, F.R.S.
Dr. H. D. KAY, C.B.E., F.R.S.
Sir BEN LOCKSPEISER, K.C.B., F.R.S.
Prof. KATHLEEN LONSDALE, F.R.S.
Prof. A. D. MACDONALD.
J. WICKHAM MURRAY, O.B.E.
Prof. F. W. PAISH.
Dr. C. F. A. PANTIN, F.R.S.
Prof. W. J. PUGH, O.B.E., F.R.S.
Dr. J. RAMSBOTTOM, O.B.E.
Prof. L. DUDLEY STAMP, C.B.E.
Dr. W. E. SWINTON.
Prof. P. E. VERNON.
Prof. W. WARDLAW, C.B.E.
W. H. WILCOCKSON.
Prof. S. W. WOOLDRIDGE, C.B.E.
Dr. N. C. WRIGHT.

EX-OFFICIO MEMBERS OF THE COUNCIL.

Past-Presidents of the Association, the President for the year, the President and Vice-Presidents for the ensuing Annual Meeting, past and present General Treasurers and General Secretaries, and the Local Treasurers and Local Secretaries for the Annual Meetings immediately past and ensuing.

PAST PRESIDENTS OF THE ASSOCIATION.

H.R.H. THE DUKE OF WINDSOR, K.G., F.R.S. (1926).
Sir HENRY DALE, O.M., G.B.E., F.R.S. (1947).
Sir HENRY TIZARD, G.C.B., F.R.S. (1948).
Sir JOHN RUSSELL, O.B.E., F.R.S. (1949).
Sir HAROLD HARTLEY, K.C.V.O., C.B.E., M.C., F.R.S. (1950).
H.R.H. The DUKE of EDINBURGH, K.G., K.T., F.R.S. (1951).
Prof. A. V. HILL, C.H., O.B.E., F.R.S. (1952).
Sir EDWARD APPLETON, G.B.E., K.C.B., F.R.S. (1953).
The Lord ADRIAN, O.M., P.R.S. (1954).

PAST GENERAL OFFICERS OF THE ASSOCIATION.

Prof. P. G. H. BOSWELL, O.B.E., F.R.S.
Prof. Sir DAVID BRUNT, Sec. R.S.
Sir HAROLD HARTLEY, K.C.V.O., C.B.E., M.C., F.R.S.
Dr. EDWARD HINDLE, F.R.S.
Sir FRANK SMITH, G.C.B., G.B.E., F.R.S.
Prof. F. J. M. STRATTON, O.B.E., F.R.S.
Sir ARTHUR TRUEMAN, K.B.E., F.R.S.

HON. AUDITORS.

Sir JOHN SIMONSEN, F.R.S. | Dr. H. J. T. ELLINGHAM.

SECRETARY.

Sir GEORGE V. ALLEN, C.B.E.

TECHNIQUES IN ARCHAEOLOGY

One of the functions of the British Association is to provide opportunities for scientists in diverse fields of work to pool their ideas. The meeting-points between different disciplines have often been most productive, and the holding of symposia on natural-science aids in archaeology has proved both interesting and fruitful.

It was at the symposium on Dating the Past, organised by Section H (Anthropology and Archaeology) at the Dundee meeting in 1947, that the application of the fluorine-dating method to the Piltdown material was proposed. At Edinburgh in 1951 one session of Section H was devoted to Ancillary Aids to Archaeology (*see Advancement of Science*, No. 33, June 1952) and at Belfast a similar symposium was held on the subject of Laboratory Techniques applied to Archaeology (*see Advancement of Science*, No. 36, March 1953). In the symposium on Techniques in Archaeology, held at Oxford on September 7, 1954, Dr. K. P. Oakley, British Museum (Natural History), described recent developments in the relative dating of bones by the fluorine method, now supplemented by nitrogen analysis and radiometric (uranium) assay, and used successfully in solving the Piltdown problem.

On the same occasion Dr. E. T. Hall, Clarendon Laboratory, Oxford, demonstrated and gave an account of his technique for X-ray spectrographic analysis of the metallic elements in the surfaces of coins, glass, glazed ceramics and the like. This method is particularly useful where adequate sampling of the specimen for chemical analysis would cause irreparable damage. It was used to detect chromium in some of the Piltdown specimens.

The contribution of Dr. G. W. Dimbleby, Forestry Department, University of Oxford, to this symposium provided a good example of the shared benefit of co-operation between scientists in different fields. Through entering archaeology, forestry specialists are beginning to obtain detailed information about the rate and character of the changes in vegetation pattern due to early cultivation. This information is not only useful to the prehistorian, but constitutes fundamental data for the plant ecologist.

Extended summaries of the three papers are printed below.

Analytical Methods of Dating Bones

by

Dr. Kenneth P. Oakley

By measuring the radioactivity of organic carbon extracted from a bone its age in years can be calculated. In specimens more than 41,000 years old the radioactivity is too weak to be measured by existing techniques. Another limitation of the C^{14} method is that at least one gram of carbon must be available for extraction.

The absolute age of a bone is more readily obtained when it has come from a deposit containing other material suitable for radiocarbon dating. Thus the palaeolithic skull found at Florisbad in the Orange Free State is estimated to be more than 41,000 years old on the basis that a sample of associated peat proved to have no measurable radioactivity.

Dating a specimen by the age of the containing deposit makes the assumption that they are contemporary, and obviously this is not always justifiable, particularly where human remains are concerned, because of man's long-established practice of burying the dead. The fluorine test (and other tests working on the same principle described below) was mainly devised for the purpose of checking the age-relation between bones and the deposit in which they have been found, or to which they are attributed.

The relative age of a fossil bone can be

determined by comparing its chemical composition with that of other fossil bones of known ages from the same site, or from the same area if they have been preserved under comparable conditions. As soon as bones are buried their composition is subject to chemical changes, some rapid, others much slower. The *organic matter* in bone consists mainly of fats and protein (collagen). The fatty matter of bone is lost quite rapidly after burial, but the protein disappears much more slowly, and under some conditions, where the soil is permanently frozen or where bacteria and air are excluded, it may persist indefinitely. Collagen fibres have been demonstrated not only in fossil ivory from Siberia, but in bone embedded in an Upper Pleistocene clay in Britain.

The appearance or texture of a bone is not a reliable guide to how much organic matter it contains. The organic content of fossil bones has generally been regarded as an unreliable criterion of their antiquity, but extensive analytical studies carried out recently have shown that some bones which appear to be well fossilised have in fact retained considerable amounts of their original protein in a denatured condition. As bones lose their protein at a slow, and under the same conditions fairly uniform, rate the relative ages of bones at some sites can be determined by comparing their organic contents—carbon, nitrogen, and chemically-bound water—or the quantity of mineral ash after burning, which becomes greater with age.

The *mineral matter* of buried bones and teeth is also subject to changes, dependent on the composition of percolating ground waters. These changes are of two kinds : (i) alteration of the phosphatic mineral of which bones are mainly composed (hydroxyapatite) ; (ii) addition of new mineral matter (e.g. lime or iron oxide) in the pores of the bone, which become open with the decay of the organic matter. Changes of this latter kind are what people normally mean by fossilization, but they can occur so rapidly and are so easily reversed (that is to say the additional mineral matter can be readily dissolved out through some local change in conditions) that they are quite unreliable as a means of relative dating. The change in mineral composition which is the most valuable criterion of relative age is the slow alteration which occurs through irreversible substitution of one element for another in the hydroxyapatite. The two elements which accumulate in this way are fluorine and uranium.

Fluorine is a poisonous gas, not found free in nature, but widely distributed as soluble fluorides which occur as traces in almost all ground waters. When fluorine ions come into contact with bones and teeth they are adsorbed, and become locked in, through replacing the hydroxyl ions in the ultramicroscopic crystals of hydroxyapatite of which skeletal materials are composed. Fluorapatite is less soluble than hydroxyapatite, so that when fluorine atoms have been fixed in bone they are not readily dissolved out, except under conditions which would destroy the whole bone. With the passage of time bones and teeth in permeable deposits accumulate fluorine progressively. The rate at which the fluorine content increases varies from place to place, but bones that have lain for the same period of time in a particular deposit will have approximately the same fluorine content. As the fluorine fixed in bone is not readily removed, a specimen which has been washed out of a more ancient deposit and redeposited in another at a later date will show a much higher fluorine content than bones contemporary with the bed. Bones artificially interred in the same deposit at a later date may have come to resemble the fossils of the bed, but will have accumulated substantially less fluorine. This is the basis of the fluorine-dating method, the usefulness of which was demonstrated in 1948 when it was applied to the Galley Hill skeleton and the Swanscombe skull.

The Galley Hill skeleton is that of a man of modern type but with certain so-called primitive traits ; it was discovered at a depth (it is said) of 8 feet in ancient river gravels containing early palaeolithic flint hand-axes and remains of extinct elephant, rhinoceros, lion, dating from before the time of Neanderthal man. In these same gravels the bones of the Swanscombe skull were found in 1935–1936 at a depth of 24 feet. The fluorine content of the Galley Hill skeleton, of the Swanscombe skull and of some twenty fossil animal bones from the same gravels was

determined by chemical analysis in the Government Chemist's Department in 1948. The results confirmed the antiquity of the Swanscombe skull, and indicated that the Galley Hill skeleton was an intrusive burial very considerably later than the gravels in which it lay.

These results have been checked, not only by redetermination of the fluorine content of the bones, using a more refined technique ; but by measurement of their organic nitrogen content. An outline of the results is shown in Table I.

TABLE I

	%F	%N
Neolithic skull, Kent .	0·3	1·9
Galley Hill skeleton .	0·5	1·6
Swanscombe skull .	1·7	nil
Bones of fossil mammals from Swanscombe gravels . . .	>1·5	traces

Although the Galley Hill skeleton is clearly much later than the gravel in which it lay, the question of how much later cannot be solved by the fluorine method of relative dating ; there are no known Late Palaeolithic, Mesolithic or Neolithic bones from immediately local deposits for comparison. Recent tests indicate that there might be sufficient original organic carbon in one of the limb-bones of this skeleton to enable its age to be determined by the radiocarbon method. If the bone were sectioned and a good cast made before its destruction in the C^{14} apparatus, the loss would be more than compensated by the gain in knowledge.

The solving of the Piltdown Problem began in 1949 with the application of the fluorine-dating method to all the specimens from the Piltdown gravel. In addition to the jaw-bone, canine or eye-tooth and fragments of human braincase, eighteen fossil mammalian specimens had been recorded from the pit. These appeared to be of two ages : an older group originally called Pliocene, including pieces of the grinding teeth of a rare type of elephant (*Elephas* cf. *planifrons*) which died out near the beginning of the Pleistocene period, that is more than 500,000 years ago according to geological estimates of an indirect kind ; and a much later group including remains of beaver, red deer and perhaps hippopotamus, dating from just before the last glaciation of about 50,000 years ago. It appeared that fossils of two formations had been washed together, through some strange chance. Those authorities who could not believe that the jawbone belonged to the skull held the view that the jaw belonged to the ' Pliocene ' group, and the skull to the later. Small samples of all these supposedly precious specimens were drilled out in the geological laboratory of the British Museum (Natural History) and submitted to the Government Chemist's Department, where they were analysed by Dr. C. R. Hoskins. The fossils of the older group proved to contain a great deal of fluorine, 2–3% (3·8 is the theoretical maximum). In striking contrast, the amount of fluorine in small samples of the famous skull, jawbone and canine tooth proved to be only $0{\cdot}2\% \pm 0{\cdot}2$ (that is to say $<0{\cdot}1$ to $0{\cdot}4\%$). It was not immediately obvious that these remains were bogus, because there did not appear to be any significant difference in fluorine content between the jawbone and skull, or between the skull and an associated molar tooth of hippopotamus (extinct in Britain since before the last glaciation).

In 1953, to test Dr. J. S. Weiner's hypothesis that the jawbone and canine tooth were modern specimens faked to match fossilized human skull bones, larger samples were removed and submitted to the Government Chemist's Department, where they were analysed by Mr. C. F. M. Fryd, using a more refined technique than that available in 1949. The jawbone and canine proved to contain no more fluorine than fresh bones and teeth, whereas the human skull bones contained just enough to indicate that they were ancient. The organic (nitrogen) content of the jawbone and teeth was also determined, and proved to be the same as in fresh specimens. The nitrogen content would not have proved the modernity of the jawbone conclusively without the cross-check provided by fluorine. A woolly rhinoceros bone from clay below the Lloyd's building in Leadenhall Street, London, has the same nitrogen content as the Piltdown jawbone, yet it is undoubtedly fossil. The reason for its remarkable preservation is that it was embedded in clay which has

provided an anaerobic and sterile environment since the latter part of the Pleistocene Ice Age. A fragment of mammoth bone from the same site, but from a layer of *sand*, contains almost no nitrogen. Fortunately the fluorine content of buried bones increases at almost the same rate in sand (or gravel) and in clay. Thus the rhino and mammoth bones both contain about 1·0% fluorine, as expected in alluvial deposits of Upper Pleistocene age in Southern England. If for the sake of argument, it were supposed that the Piltdown jawbone had been encased in clay since Pleistocene times, and this had prevented its organic content from being degraded, its fluorine content should nevertheless be higher than in modern bone. In fact its matrix was said to be gravel.

The fluorine content of the Piltdown skull bones is lower than that of any fossil bones from Pleistocene gravels in Britain. But so long as the hippopotamus tooth with the same low fluorine content was accepted as a local fossil it could be argued that the ground-water at Piltdown had been exceptionally deficient in fluorine since the Ice Age. (The exposed dentine of a tooth and compact bone adsorb fluorine at about the same rate.) When evidence came to light that the hippopotamus tooth and, indeed, almost all the Piltdown fossils had been artificially stained, and therefore were fraudulent introductions, the provisional dating of the skull as late Pleistocene lost its foundation.[1]

Further chemical studies of the Piltdown hippo molar showed that its organic content was almost negligible, indicating that it was a fossil of considerable antiquity—yet its low fluorine content would have suggested, in a gravel matrix, that it was *not* very ancient. It has now been found that in limestone cave desposits, passage of fluorine ions is inhibited, but that decay of organic matter may proceed at normal rates. Hippopotamus remains are rare in cave deposits, except in Mediterranean islands, for example Malta, which is a possible source of the Piltdown specimen.

A new method of relative dating of bones and teeth has been tried with considerable success on the Piltdown specimens : the method of *radiometric assay*. It was established by Lord Rayleigh in 1908 that mineral phosphates, including fossil bones, contain uranium. Medical research has since shown that uranium circulating in the blood-stream is fixed in the phosphatic mineral matter of bones by replacement apparently of the calcium atoms in the hydroxyapatite. The same process of replacement occurs in bones buried in deposits through which there is a percolation of ground-water containing traces of uranium. The longer a bone has lain in that deposit the more uranium it will have absorbed. Uranium is radioactive, and consequently it is possible to estimate the uranium content of a fossil by counting, under suitable conditions, the rate of its beta radiations. Radiometric assays of fossil bones carried out in the Atomic Energy Division of the Geological Survey by Dr. C. F. Davidson and Mr. S. H. U. Bowie have shown that although there is wide variation in the radioactivity of fossils of the same age from different sites, there is a progressive build-up in the average radioactivity of fossils with increasing geological age. Moreover, contemporaneous bones and teeth from a single site show little fluctuation in radioactivity, although there are some exceptions. (The radioactivity of the enamel of a tooth is considerably less than that of dentine which is more adsorptive ; while fossil bones and teeth from limestone formations and clays contain less uranium than specimens from gravels and sands of the same age.) The radiometric assay can serve in the same way as fluorine analysis, to distinguish older derived specimens from younger contemporary or intruded specimens, when remains of more than one age occur together in the same deposit. It has an advantage over the fluorine method of relative dating in that it does not involve destruction of material.

When Bowie and Davidson carried out a radiometric assay of all the Piltdown specimens, the results agreed with those obtained by fluorine analysis, indicating that the skull is more probably post-Pleistocene than Pleistocene and that the animal remains have been derived from very varied geological sources. The most striking additional information concerned

[1] Eventually it was proved that the skull bones had been transplanted and therefore were of uncertain origin.

TABLE II

	%F	%N	eU_3O_8 p.p.m.[2]
Fresh bone	0·03	4·0	0
Piltdown jawbone	<0·03	3·9	0
Neolithic skull, Kent	0·3	1·9	1
Piltdown skull	0·1	1·4	2
Piltdown hippo molar[1]	<0·05	0·06	3
Malta hippo molar[1]	0·1	nil	7
Swanscombe skull	1·7	nil	27
Suffolk Red Crag, *Mastodon* molar[1]	1·9	trace	46
Piltdown *E.* cf. *planifrons* molar[1]	2·7	nil	610
Ichkeul *Elephas* cf. *planifrons* molar[1]	2·7	trace	580

[1] Samples of dentine or cementum.

[2] See explanation on p. 6 and in Table III.

the fragments of elephant tooth comparable with molars of *Elephas planifrons*, which lived in the earliest stage of the Pleistocene (Villafranchian), and has been rarely if ever recorded elsewhere in Britain. These fragments proved to have a radioactivity far higher than that of any Tertiary of Pleistocene fossils from Britain that have been tested, and higher than that of fossil elephant teeth of the same age from foreign localities, with the only exception of a specimen of *Elephas* cf. *planifrons* from Ichkeul in Tunisia (Table II).

There are so many variables affecting the composition of fossil bones and teeth that it is usually impossible to use either fluorine content or radioactivity as more than a rough guide to the geological age of an isolated specimen. However, these methods, combined where necessary with nitrogen analysis, are useful in (1) establishing the relative ages of a variety of vertebrate remains occurring in comparable circumstances at the same or neighbouring sites, and (2) tracing the most likely origin of specimens derived from several possible sources. The usefulness of the combined techniques is well illustrated by the results obtained through applying them to the problem of the relative age and origin of sabre-tooth tiger remains reported from British cave deposits.

In 1876 a canine tooth of *Machairodus* was found in an Upper Palaeolithic layer at Robin Hood's Cave, Creswell Crags, Derbyshire. Some authorities have doubted whether this was a genuine find, suspecting that it had been fraudulently planted at the site, and that it had really originated in alluvial deposits of earliest Pleistocene (Villafranchian) age in France or Italy, where specimens of this genus are not uncommon. Comparison of the fluorine content of the Creswell tooth with that of Villafranchian specimens from the main Continental localities appears to dispel this suggestion (Table III). Other possibilities have to be considered. Was the Creswell canine derived from an early Pleistocene deposit in Derbyshire and brought into the Robin Hood's Cave by prehistoric man? There are several instances of fossils having been treasured in the Stone Age and transported many miles from their natural source (e.g. the Silurian trilobite found in a Magdalenian layer in Grotte du Trilobite at Arcy-sur-Cure (Yonne)). *Machairodus* canines (but specifically different from the Creswell

TABLE III

Source of *Machairodus* dentine tested	%N	%F	$\%F/\%P_2O_5$ (× 100)	eU_3O_8 p.p.m.[3]
Val d'Arno, Italy	0·2	1·6	5·3	34
Mt. Perrier, France	nil	1·9	6·3	26
Doveholes, Derbyshire	nil	2·5	8·6	68
Creswell, Derbyshire	2·1	0·2	0·8	<1
Kent's Cavern, Devon	1·2	<0·01	<0·1	5

[3] Uranium content estimated on basis of net beta radiations per minute.

specimen) have been found in association with Villafranchian fossils in a fissure deposit at Doveholes, also in Derbyshire. In their lack of organic matter, high fluorine content and high radioactivity they contrast sharply with the Creswell specimen, whose composition agrees with that of local Upper Pleistocene cave mammals.

If it were still to be maintained that the Creswell canine was fraudulently planted in the cave, its composition could only be accounted for by supposing that it had been obtained from some other similar cave deposit elsewhere ; for in our experience it is only in limestone cave deposits that Pleistocene vertebrate specimens are so deficient in fluorine as this one. Teeth of *Machairodus* (a genus confined in Pleistocene times to the Old World) have been reported from Upper Pleistocene (Mousterian) cave-earth in Kent's Cavern, Torquay. It has generally been assumed by vertebrate palaeontologists in recent years that these specimens were residues from some much older deposit in the Torquay Cave system. The composition of one of the Kent's Cavern canines (BMNH. Geol. 14954) has now been tested. The fluorine content of the dentine is negligible, and the uranium content is low ; but the failure of these elements to circulate in calcareous cave deposits is now recognised. However, if these specimens were considerably older than the undoubtedly Upper Pleistocene mammoth teeth found in the same cave-earth they should contain substantially less organic matter. In fact they prove to contain rather more (Table IV). Thus it may be concluded provisionally that sabre-tooth tigers survived in Britain until Upper Pleistocene times.

TABLE IV

Specimens of dentine from Kent's Cavern	% N	% F/% P_2O_5 (× 100)
Machairodus canine ; cave-earth	1·2*	<0·1
Mammoth molar ; cave-earth	0·8	<0·1
Bear molar ; basal breccia	0·2	<0·1

* Washing the sample in warm water and then in acetone effected no reduction in nitrogen content.

THE SPECTROSCOPIC METHODS OF ANALYSIS IN ARCHAEOLOGY

by

Dr. E. T. Hall

ARCHAEOLOGISTS have for a long time wanted to find out more about what they handle, but museum curators have been loath to let scientists touch their precious exhibits. Unfortunately a method of analysis where absolutely no damage is caused is of no particular interest except to archaeologists, and so very little research has been done on the subject, and there was no suitable method to borrow from other fields. It was for this reason that the X-ray spectroscope was developed.

Why do we want to know what our coins, for instance, contain ? This can be put under two headings. Firstly, from the point of view of the ancient historian, an analysis of a number (in some instances a matter of hundreds) of specimens from different sources may give us evidence of the existence of trade between two towns or countries. If we can establish the existence of an alloy which is repeated elsewhere, we have at least some evidence for interchange of commerce between the two areas. A large number of specimens would have to be examined because a form of statistical investigation would be necessary to get any definite results. What we need is a whole series of results so that some really concrete conclusion can be obtained. For this reason a vast amount of work will be necessary to build up the background from which these conclusions can be drawn. A whole library of tabulated results is necessary for comparisons to be made.

The second type of problem, in which a knowledge of composition can help us, is where we want to know something about the processes and techniques used in making some particular object.

Finally, perhaps the most amusing revelations which an analysis can show, are those where we are concerned with the genuineness of a specimen, whether it be a wrong attribution or a deliberate fake.

When we are looking for a method of analysis there are some essential points to be borne in mind. Firstly, as far as possible the object must not be marked or damaged in any way, and secondly, the method must

be as rapid as possible. Detection of trace impurities down to and below 1/100 of one per cent. would be helpful. This is particularly desirable when the metals are of high purity, as is the case with much of the gold found in the ancient world. Incidentally, an absolute accuracy would not be of such great importance—in many cases dependable analyses which could merely be compared with the other, would be sufficient.

A non-spectrographic method which has been used on the experimental scale is neutron bombardment in a uranium pile. Of course, the main disadvantage is having to take your specimen to a suitable source of neutrons, which is at present only available at Harwell. The process however, is very simple. One places the coin (the method is particularly suitable for this type of object) in front of one of the beams of thermal neutrons emerging from the pile and leaves it there for several hours or days, during which time a certain portion of the atoms in the coin will have become radioactive. If we now take the coin away from the pile and study how this radioactivity decays with time, we can in theory find out what elements are present in the specimen, and to what extent. This is because some radioactive atoms decay, to what they were before, quicker than others. Unfortunately this type of investigation covers a long period, although a large number of specimens can be analysed at the same time.

It is possible, however, that in special cases where the X-ray spectrometer is not entirely suitable this method might have possibilities. It has been found that ½ per cent. of platinum in gold may be identified in this way, whereas with the spectrometer these two elements are difficult to separate.

We now come to a method of analysis which has found a certain amount of favour with the museum authorities—this is the ordinary optical spectroscope. In this apparatus a spark is struck between the specimen to be analysed and another pure electrode. The light produced in this way is then split up into its component parts by means of a prism or diffraction grating. The arc struck in this way is extremely hot, in fact about 4000° C., which will be sufficient to melt the area immediately around the arc. Furthermore, the velocity of the atoms which are accelerated by the electric field will be of the order of a million cm/sec. The combination of these effects will be to cause pitting of the surface of the object—although by keeping the arc as small as possible this pitting can be kept to a minimum. Nevertheless, on small objects such as coins, the damage is sufficient to prevent the examination of all coins except those of little importance. However, for large objects this way of analysing is satisfactory, particularly as the method is quite fast. A photographic record of the spectrum and the measurement of, perhaps, four different elements on a microphotometer, followed by their interpretation in the form of percentages can, in suitable circumstances, be done in fifteen minutes.

The sensitivity of the method is remarkable. Percentages down to 1/1000 of one per cent. are detectable for nearly all elements. However, perhaps the most serious drawback, besides the damage question, is the fact that it is not suitable for proportions in excess of 4 per cent. If, for instance, we are interested in the tin content of bronzes, or, perhaps, the composition of electrum, we are not able to use the optical spectrometer for quantitative work.

It seemed, therefore, when this investigation started, that some different method of analysis was desirable and it appeared that the possibilities of X-ray analysis had not been fully investigated. The fluorescent method of analysis was used, where the specimen is situated outside the X-ray tube, and rays from a standard tube excite rays characteristic of the elements in the sample.

We have to be able to distinguish one variety of characteristic radiation from another, and the spectrometer is used for this purpose. When X-rays strike a crystal in the ordinary way, they are scattered or pass straight through, but if that crystal happens to be at exactly the correct angle to the incident radiation for a certain wavelength, the rays will be reflected as though the crystal was a mirror. This angle is known as the Bragg angle for that combination of crystal and X-ray wavelength. By using a Geiger counter which will register an X-ray if it passes through, we can measure the number per minute being reflected. In this way we have an indication of the element (by the angle of

the crystal) and also the amount of that element (by the number of counts in unit time).

If we have no idea of the contents of what we are examining then we must scan through all the possible angles of the crystal, counting for a set amount of time in each position. This is very tedious, so the scanning process is made automatic. The apparatus counts for a pre-set time, then stops counting while the spectrometer is stepped on a certain number of minutes of arc, and meanwhile records that particular number of arcs on a calculating machine, which also subtracts any cosmic background. Then the recording mechanism is reset to zero and a new count is made in the new position of the spectrometer.

In this way we can construct a curve of angle against intensity. The height of the peaks which will occur where there is a particular element present will give the percentage of that element in the sample. These graphs are constructed on an automatic graph drawer. In the case of small objects such as coins, we can place up to eight different specimens which are analysed one after the other. By scanning through the whole useable spectrum, we can analyse for all the elements between potassium and uranium in the Periodic Table. Unfortunately, the characteristic radiation of the lighter elements is too soft, and is heavily absorbed by the air on the way from the specimen to the counter, and without special precautions they cannot be estimated. To cover all these elements from potassium upwards is rather slow ; about eight hours being necessary to get a sensitivity of one-tenth of one per cent. However, if we are only interested in certain elements, there are facilities in the apparatus by which one need only count at the required points in the spectrum.

The spectrometer will only count at those points which correspond to the elements required. In this way we can estimate each element in about five minutes.

Let us now consider a few of the problems which have been investigated to show the wide scope of non-destructive analysis.

(1) *Athens and Corinth*

An example is afforded by Athens, which had, perhaps, the largest silver mines in the Mediterranean, at Laurium ; moreover, her people took great pride in producing really high quality silver. In fact the purity of this silver is truly remarkable, and many of the coins analysed were over 99 per cent. pure silver. Corinth on the other hand had, so far as we know, no such silver deposits, and the question arises as to whether she obtained her silver for coinage purposes from Athens or elsewhere. In spite of being neighbours they were frequently not on friendly terms, and so quite possibly Corinth went further afield for her supplies. It was found in fact that impurities in the two series of silver differed considerably, and so the supposition that they did not go to Athens for their silver seems to be the correct one.

(2) *Damastium and Chalcis*

Damastium is believed to be a town in Northern Greece, and it is known that she had large silver mines, and so it is reasonable to assume that she struck her coins from local silver. At the same time the influence of Chalcis was very strong, but she did not have any silver of her own. Did she obtain her coinage requirements from Damastium ? It is also possible that these mines were one of the sources from which Corinth obtained her silver. All the coins analysed which came from Damastium show a relatively high cadmium content ; this is most unusual and should provide a means of tracing Damastian silver from anywhere. This cadmium content is not found in either the coins from Chalcis or Corinth.

(3) Another old problem is posed by the Greek electrum coins. Electrum is the name given to the variable mixture of native silver and gold. Many ' analyses ' have been carried out on these coins by measuring their densities, and then assuming they consist only of silver and gold, one can in theory calculate the percentage of the two metals. However, this method gives little idea of their composition, and erroneous conclusions have been drawn. For instance, it has been thought that all electrum was obtained from the natural

alloy, and the composition varied according to the metal as it was found in nature. In fact, we find that the gold content is remarkably constant series by series, and in all probability the mints took quite a bit of trouble to match their electrum so that each coin had approximately the same value. Quite often we get electrum and gold coinages of the same state being issued simultaneously, and it would seem peculiar if they did not keep the composition constant. We sometimes have instances where the electrum composition is so constant that the only explanation would seem to be that they purified the metals separately and then remixed them in certain proportions. For instance, in the Lesbian electrum series this is particularly evident, and, incidentally, one finds considerable quantities of copper added to these alloys—perhaps a cunning device to give the alloy a more 'golden' appearance, and so enhance its apparent value.

The great advantage of X-ray analysis is the fact that it is immaterial what the chemical state of the element is. For instance, the composition of china, porcelain or glass, can easily be determined by X-ray spectroscopy. The problem chosen here to demonstrate this is the attribution of dates to eighteenth-century English porcelain. Many of the early pieces were not marked, and it is difficult to know sometimes whether a sample is of the earliest period or not.

We are lucky in the fact that at a certain period the composition used in making the porcelain changed. The old method in the mid-eighteenth century was to heat a mixture of sand, soda and crushed lead glass. It was found, however, that since there was no accurate control of the temperature of the kilns, the structure of the porcelain would very often collapse. It was the factory at Bow which discovered that if one replaced the lead glass in the formula with bone-ash (that is to say calcinated bones) one obtained a much greater rigidity. Since we know that the change over from one formula to the other took place at certain dates in each factory, we can by analysis, say whether a certain piece was made before that date or not. For instance, the change over took place in 1750 at Bow and Chelsea, and in 1770 in Derby.

Lastly, might be mentioned the use of the X-ray spectrometer in identifying the chromate and iron staining of the Piltdown fragments and flints. Here we have a typical example where this type of non-destructive investigation can be of use on unique specimens. Moreover, since the staining, particularly of the flint, was extremely thin, but covered a comparatively large surface, the method was extremely sensitive in the estimation of the chromium present.

The Ecological Study of Buried Soils

by

Dr. G. W. Dimbleby

The construction of an earthwork consists essentially of putting a mass of material upon a firm foundation. As a result, that foundation is in effect preserved and is no longer subject to those changes which are associated with a surface ; the top of the earthwork is the new surface. Of course, certain changes do take place at the foundation level with the passage of time, but it has been found that this buried surface can yield interesting information concerning conditions at the time it was buried. In some cases an examination of the bulk of the earthwork itself may also repay study, as we shall see, but it is the buried surface upon which I want to concentrate.

This surface may either be the original surface upon which vegetation had been growing or it may be a prepared surface. In the latter case most of the material of value for our purpose has been removed since it is the top inch or two of a soil which tells us the most about the land use at the time. However, the absence of a well-defined humus layer at the base of an earthwork does not necessarily imply that the old surface has been destroyed. Under certain conditions the humus can gradually disappear even when buried. Nevertheless, one is much more confident when some trace of a humus layer can be seen. Very often under barrows the old foundation is also indicated by a thin line of

powdered charcoal, though this could appear equally well on a prepared surface.

The visual appearance of the buried soil may give useful information as to the general type of vegetation present when the earthwork was built. If the humus is dark and concentrated in the upper layers of the mineral soil, grassland may well have prevailed. A dark band of humus, often greasy and peaty, overlying a bleached sandy layer is typical of the podzol which develops under heather. Finally, where no humus is visible at all, whilst this may indicate a prepared surface, it might also be associated with primitive cultivation, particularly on poor sandy soils. These conclusions cannot stand by themselves, but taken together with other features they help to complete the picture.

Another feature of the visible profile which is significant is the presence or absence of stones near the surface. Where the underlying geological material is a stone-free sand no conclusions can, of course, be drawn, but in other cases the existence of several inches of stone-free soil at the surface indicates prolonged earthworm activity under undisturbed vegetation. Such a layer is found in old deciduous forest, or under grassland which has not been ploughed up for centuries. On the other hand, the presence of stones right up to the surface probably indicates exposure of the soil, allowing frost to get into the soil in the winter. As most farmers on stony land know, the stones keep coming up as fast as they are removed. Such a condition, therefore, may be a strong pointer to cultivation at the time the earthwork was built, though alone it could hardly be accepted as proof. A section of the bank of a disc barrow excavated by Mr. Rhodes at Grafton, Wiltshire, showed the present grassland soil almost free from chalk fragments ; there was no evidence of this land having been anything else but pasture in historical times. Below this was a layer of chalk rubble, thrown up from the ditch, burying beneath it another pronounced humus layer, almost certainly the old soil. This, however, was not so markedly stone-free, a point which supported the conclusion suggested by other evidence that this surface had been cultivated before the barrow was built.

The visual appearance of a buried soil may tell us at the most that the land was under grassland, heathland, deciduous forest or some form of cultivation. It cannot give us any details. These may be provided by plant remains in the soil. Charcoal is an obvious form in which these may occur and it is standard archaeological practice to collect such material. Usually it is in small discrete fragments and there is no proof that these have originated from trees or shrubs growing on the site. Occasionally, however, one finds what is obviously the remains of a root in situ and that is a little more conclusive. There is no good evidence to my knowledge that wood can be preserved in terrestrial soils unless it has been charred ; even in the most acid soils, wood which has not been burnt rots and disappears relatively rapidly.

Perhaps the most useful source of information about the ecological conditions at the time of the construction of an earthwork is provided by the pollen which may be preserved in the old soil surface. It is unfortunate for those who work in chalk or limestone regions that alkaline soils do not preserve pollen; on the more acid soils, however, pollen may be found in great quantity. Fortunately the pollen of most of our tree species is easily identified, whilst the Gramineae, the grass family, and the Ericaceae, the heather family, have very characteristic pollen. The actual counting and identification of pollen grains can become routine, but the interpretation of the figures cannot be precise. We have a fair idea of which species are over-represented and which are under-represented by their pollen, but some pollen will also be blown on to a site from vegetation some distance away, vegetation which may be quite different from that on the site. One can only reliably deal, therefore, with those species which are represented in considerable quantity and which were apparently growing on the site.

Most of my field work has been carried out on the North Yorkshire Moors, one of the most extensive areas of heathland in the country. Round barrows are frequent but there is little evidence of Neolithic occupation of these heathlands. A problem which had to be solved was whether the heathland vegetation was primaeval or secondary ; if secondary,

what preceded it? The particular area I was dealing with was devoid of deep peat so I could not use normal pollen analytical methods. So I cut into the wall of a round barrow, carefully avoiding the centre, and prepared a section down to and below the old surface. This, it so happened, was well marked by a layer of charcoal, all of which proved to be oak and birch—point number one. Point number two was most striking: the present podzol soil did not extend under the mound but went over the top. It was therefore secondary, and since this soil type is closely associated with heath, it implied that the heath was also secondary. The soil which underlay the barrow consisted of orange sand right up to the old surface; it was not visibly leached. One assumes, therefore, that the soil was of the brown forest type, typical of deciduous forest, but also found with variations under pasture or cultivation. However, no concentration of humus was apparent, and whilst it is quite possible that this would have disappeared by now in any case, the probability was that the surface was either forested, or under cultivation, or, for that matter, a prepared one. The absence of a stone-free layer accorded with the possibility that the surface had either been under cultivation or was prepared. It is worth noting that the main mass of the mound was also made up of orange sand, suggesting that this was available in quantity nearby, a further pointer to the absence of the leaching influence of heather in the immediate district.

Microscopical examination of samples from the old surface proved rich in pollen, so it was obviously not a prepared surface. The analysis of this pollen is shown in Table I.

Most striking is the small proportion of pollen from the Ericaceae, the heather family. Under present conditions this makes up about 90 per cent. of the pollen, but here the percentage is less than 4. There is a good assortment of deciduous trees though not in such proportions as to indicate closed forest. The high percentage of hazel, suggests that high forest had given way to hazel scrub. But the herbs are significant because amongst them there are several weeds of cultivated land which would not occur either in forest or hazel thicket. One can therefore build up a picture of a patch of cultivation in a landscape well covered with hazel thicket, but with patches of high forest still remaining here and there, an incredible contrast with the monotony of the present-day heath.

Two local archaeologists, W. H. Lamplough and J. R. Lidster, have their spare time fully occupied in dealing with rescue digs, since the area is under intensive afforestation, but as the barrow I have just been dealing with is not threatened, it has not been excavated and therefore remains undated. I was fortunate, however, in being able to join Lamplough and Lidster in a rescue dig which provided me with further interesting material. It was a small round barrow (Springwood, No. 22); the primary cremation revealed a pygmy vessel in the form of a miniature urn and the base and lower parts of a larger vessel. Above this cremation a secondary cremation was found, much disturbed by more recent excavation, with associated sherds of an overhanging-rim urn decorated with cord impression both on and below the rim, but devoid of neck or shoulder. Comparing it with other local material, Lamplough suggests that it belongs to the later half of the Urn period. The fact that both cremations were of the Middle Bronze Age B gives a maximum time interval according to Lamplough of

TABLE I

Number of Pollen Grains and Spores counted from Floor of Tumulus (Thieves' Dykes)

Trees		Herbs	
Pinus	1	Gramineae	15
Betula	9	Cyperaceae	6
Alnus	17	Scleranthus type	14
Quercus	29	Chenopodiaceae	5
Tilia	10	Plantago	3
Ulmus	1	Succisa pratensis	23
Acer	5	Compositae	*
Shrubs		**Pteridophytes**	
Ericaceae	14	Dryopteris type	60
Corylus	69	Lycopodium type	9
Salix type	2	Polypodium	112
Sorbus type	3	Trilete spores	16
Ilex	*	Total	423

* Recorded in preliminary scan only.

roughly 400 years, with a strong probability that it was less than half this. When the second cremation was deposited in the barrow, the whole mound was covered over with soil to a depth of 3 or 4 inches and a new ditch was dug around the perimeter. The primary and secondary ditches had each been filled up with heather raw humus. The original surface on which the mound was built was clearly visible in section, and the primary mound surface was still outlined in places by a humus streak.

were taken on a profile from the present surface to the subsoil alongside the primary cremation, on another face just through the original surface, and finally through the double ditch.

Fig. 1a shows selected groups from the first two series of pollen profiles. In the zone of burnt soil, pollen was largely destroyed and from one or two layers satisfactory counts could not be obtained, whilst the adjacent analyses must also be regarded as possibly unreliable. Though this affects the original surface, the

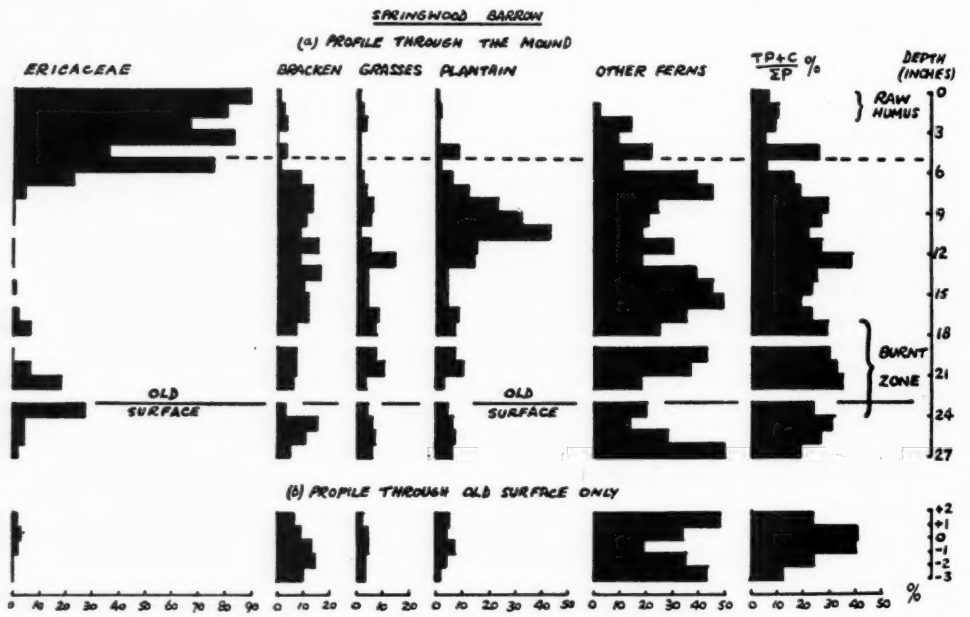


Fig. 1 (a) and (b). $\frac{TP + C}{\Sigma P} = \frac{\text{Tree pollen + hazel pollen}}{\text{Total of all pollen}}$

Visual observation of the section showed that leaching of the soil under the barrow had just started. There were a few inches of greyish soil merging directly into the orange sand beneath. It was also noticeable that the main bulk of the mound was more leached than that of the first barrow. Some of this leaching may have taken place since the mound was built, for the material spread over and to one side of the primary cremation was reddened by burning, which could not have happened if the iron had been leached from it first.

Charcoal was found in abundance in the mound and without exception proved to be oak. Samples for pollen analysis

duplicate sampling through this surface, shown in the lower diagram (Fig. 1b), was taken from a place where there had been no burning, and revealed a very low percentage of Ericaceae. This appeared somewhat higher at the old surface in the long series. Compared with the upper layers of the profile, however, it can be seen that both these amounts are small, and the absence of heather pollen in the wall of the mound, which was probably made up of topsoil, means that there was no continuous expanse of heath. The proportion of woody species was much higher then, likewise bracken and other ferns. Perhaps the most startling feature of the profile is

the big peak in plantain pollen at a depth of 6–12 inches below the present surface. This mound consists of material heaped up, presumably layer by layer, and it is reasonable to expect that the pollen composition will vary according to the place where the soil was collected. Whilst downwash through the mound since its construction has doubtless blurred the picture, this effect can be shown to be negligible for present purposes. It can be inferred, therefore, that soil used to build this barrow came from land with a high weed population. Even in the old surface the plantain figure is considerable and must imply cultivation. The sporadic occurrence of cereal grains confirms this conclusion.

To summarise the ecological information provided by this investigation, we may say that the barrow appears to be somewhat later than the first one investigated, being built when forest destruction had proceeded further, particularly in favour of cultivation. The barrow was built on or among plots of cultivation which had probably been abandoned. These had been occupied by a dense weed flora which was probably replaced by bracken. This, by increasing the fire risk, and by virtue of its own litter properties, could have been responsible for the initial leaching of an already impoverished soil, a condition which would soon favour heather. At any rate, in the course of the Middle

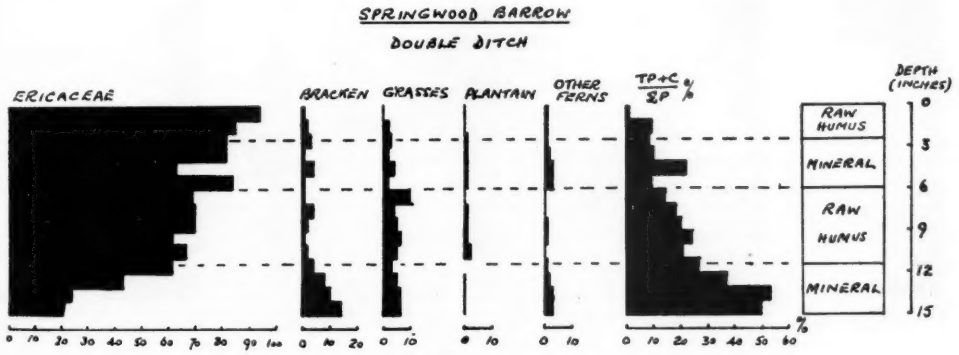


Fig. 2.

At the top of the profile it is possible to detect the level of the first surface of the mound contemporaneous with the primary cremation, even though at the point of sampling this layer was not visible. The percentage of Ericaceae at this level (76 per cent.) should be compared with the small values at the old surface, for this represents the development of the heath flora in the interval between the primary and the secondary cremations. Similar values are shown from the raw humus of the primary ditch ; though less than the percentages in the newer ditch, they are consistently around 70 per cent. The most striking difference, however, between these two raw humus samples (Fig. 2) is in the proportion of woody species ; this is much higher in the primary than in the secondary ditch and illustrates the further retreat of woody species since the secondary cremation.

Bronze Age B alone there seems to have been an almost explosive development of heathland with the consequent reduction of deciduous woodland, a process which has continued though more slowly to the present day.

Parallel trends may be seen in other parts of the country. In the New Forest and Dorset similar examples of the replacement of forest by open country can be demonstrated, whilst evidence has been obtained there also that the process became almost explosive at one stage, namely, during the time of the Wessex culture. In fact, it has proved possible to ascribe a rough dating to otherwise undatable material on the basis of the degree of deforestation revealed by the pollen analyses in comparison with those of other dated material. One difference in this region as compared with Yorkshire is that even the Early Bronze Age barrows overlie

a podzol soil. This is very likely due to continued interference with the vegetation during Neolithic times ; though there were no Neolithic settlements on these soils, there is no doubt that hunting was carried out and hunters frequently use fire to open up forested areas. Many finds of Neolithic material have been recorded, particularly on what are now the Dorset heaths, and this contrasts with the North Yorkshire Moors. Here, it seems, the first down the face of an open sewerage trench at a point which might just have intersected the old surface (Fig. 3), I was probably dealing with a ditch filling only. It is interesting to note that the lower half of the diagram shows a much higher proportion of woody species, but evidence of cultivation extends right down to the bottom in the form of weed and cereal pollen. There is a suggestion that at one stage cultivation lapsed, followed by a big

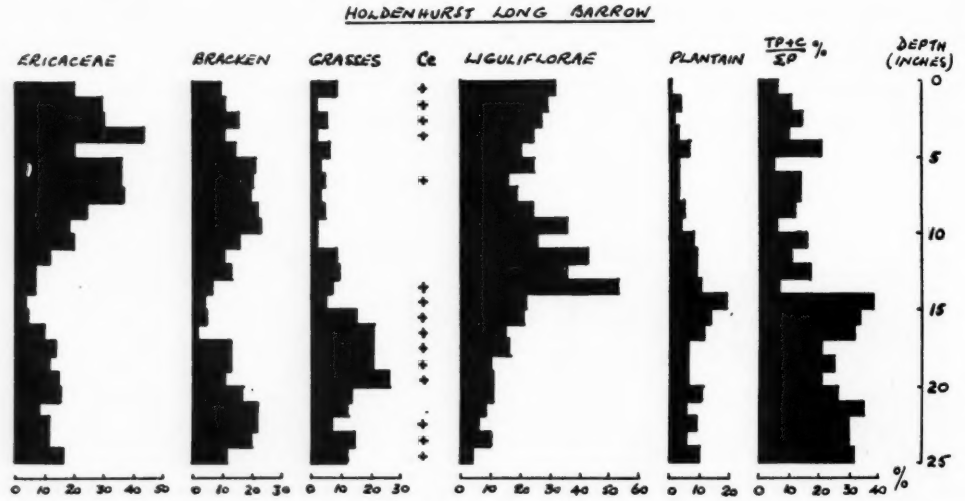


Fig. 3. Ce = cereal pollen.

Bronze Age settlers found virgin forest in association with a brown forest soil.

I have tried to take this matter back to its logical conclusion in the South by trying to find a Neolithic monument on land that is now under heath. So far, I have not found a satisfactory site. The nearest to my requirement was the long barrow at Holdenhurst (Hampshire), which, although not on heathland to-day, is at least on moderately acid gravel. My visit there just post-dated the erection of some bungalows on the site, and though I took a long series of samples spread of weeds which eventually gave way to bracken and heather. Whilst not laying too much stress on these figures, it is interesting that they point to the same ecological sequence that has been deduced elsewhere.

My own interest in this subject must remain botanical rather than archaeological, and I shall hope to find sites less distant in time from the virgin forest than those I have so far studied. By this means we should get a clearer picture of the natural forest on acid soils, a subject about which ecologists still know very little.

ANIMAL BEHAVIOUR

Sections D (Zoology) and J (Psychology) devoted the morning of September 6, 1954, at the Oxford Meeting of the British Association, to a joint discussion of animal behaviour. Dr. N. Tinbergen, Oxford University, reviewed certain problems of behaviour that can be more easily observed in animals than in man and whose study might help in elucidating human behaviour. Dr. R. A. Hinde, Cambridge University, then surveyed the modifications of inborn behaviour due to learning processes and the nature of changes in responsiveness to a constant stimulus. The symposium was concluded by Prof. A. Rex Knight, Aberdeen University, with a discussion of the implications for psychologists of the two previous papers.

These papers are printed below in the order in which they were delivered.

Some Aspects of Ethology, the Biological Study of Animal Behaviour

by

Dr. N. Tinbergen

Introduction

First of all, I want to express my gratitude to the initiators of this joint meeting for providing this much-needed opportunity for closer contact between psychologists and ethologists. We zoologists studying animal behaviour know that we have still quite a long way to go, and that we can learn much from colleagues in other, more developed sciences. But we are also convinced—and in this we are encouraged by the comments made by psychologists, psychiatrists, psycho-analysts and zoologists—that our approach and our results may be helpful to psychology because they call attention to phenomena that have so far escaped attention, or at least have been neglected.

Although ethology is still in its infancy, it is already much too extensive to be covered in half an hour. Even though Dr. Hinde and I have agreed, for this meeting, on a certain division of labour, each of us had to make a choice from the many items that deserve discussion. In view of the recent publication of a sharp criticism of ethology by D. Lehrman (2), the gist of which will be summarised by Professor Knight, it is tempting to enter into some of the points raised by this paper. While, as you will notice, I am inclined to accept part of Lehrman's criticism, I am also convinced that much in it is based on misunderstanding, on a failure to see the wide diversity of phenomena to be observed in the multitude of animal forms, and, partly, even on ignorance of facts.

Recent ethological work is often referred to as 'Lorenz's new theory of animal behaviour.' He himself would be the last man to give his work this title. It is true that he formulated several new hypotheses, but those really informed about recent developments know that these are all very much in the melting pot. However, what is new in ethology is the type of approach. Three aspects deserve special mention. (*a*) There is a deliberate return to preliminary description of observable behaviour as an indispensable basis for the formulation of problems. To us, impressed as we are by the huge diversity of behaviour phenomena, it seems a little rash to base psychological theories on what is known about Man and a few other mammals, notably the domesticated rat. (*b*) There is emphasis on the need of work on total behaviour patterns. (*c*) Ethology aims at a balanced study centred round the three major biological problems: causation, survival value, and evolution. Psychology seems to neglect the two latter problems unduly.

In spite of all its shortcomings, ethology may help in expanding psychology, and so will, it is hoped, ultimately merge with

present-day psychologies into one science of behaviour.

Non-learnt Aspects of Behaviour

The repertoire of movements of each animal, even of the most versatile mammal, is limited. The occurrence of intensity scales, combinations of patterns or parts of patterns, and of orientation components accounts for great variety, but the variability is not endless. For instance, a gull never retrieves an egg by using the feet or the wing, both of which are better suited to it than the bill—which it invariably uses. Simple, but detailed and repeated observation reveals that it is relatively easy to find out something about the factors that start off a certain type of behaviour, and about the final result, the motor pattern. Other variables being constant, a rigid connection is found to exist between ' input ' and ultimate response, and simple experimental analysis can tell us something about the achievements of whatever system is responsible for this. As an example of such an analysis I refer to work on the food-begging response of gull chicks (7). For the herring gull it has been found that this visual response is released by only a minor part of the total information available to the chick's eyes. Thus the colour of the red patch on the lower mandible, its contrast with the yellow colour of the rest of the bill, the bill's position, its lowness and its nearness are affecting the response, whereas head colour and head shape are not. Similar results have been obtained in many other species, invertebrate as well as vertebrate. Different responses of the same animal are usually dependent on different sets of such ' sign stimuli,' even though the responses may be released by the same object (5). In gull chicks, this selective responsiveness has been shown to be present before conditioning can have taken place. Many of these sign stimuli can be described only in terms of relationships ; they are configurational.

Lehrman objects to our calling such responses innate. He argues that the ontogenetic development of any character, bodily or behavioural, is a complicated series of events in which internal and environmental influences interact. I have to admit that our application of the term ' innate ' to a character (instead of, as is usual in genetics, to differences) is misleading, and that we will have to formulate our findings in another, more cautious way. What remains is that, between hatching and the time our tests were made, the gull chick had not been conditioned to the stimulus situation to which it was found to respond, and I maintain that this is a distinct contribution to the study of the ontogeny of the begging response. This type of fact has long been overlooked by psychology, mainly (I am convinced) because psychologists have shown a preoccupation with problems of learning.

A similar lack of interest in the achievements of mechanisms prior to learning strikes one in work done on motor patterns. Here again ethology has called attention to an aspect of ontogeny that had so far received insufficient attention. The type of learning usually thought of as affecting motor patterns is ' practice '—imitation being extremely rare among animals. Grohmann (1) has shown that the improvement of flight in pigeons between hatching and fledging is not due to performance of flight movements, but can be observed in pigeons that have been prevented from ' practising.' Our sweeping conclusion used to be : flight is an innate motor pattern (although it must be recognised that we often amended this shorthand statement by pointing to subsequent development of skill). If we formulate our conclusion in a more cautious way, we specify the response, the species, the period of development studied, and the learning type that has been eliminated, and then Grohmann's work gives us information, which again is a distinct contribution, however small, to the huge complex of problems concerning ontogeny.

Other approaches are possible : Weiss's tendon-crossing experiments (8) in Axolotls and rats, and clinical observations on Man, show that readjustment is impossible in these two animals, and possible to a very limited extent only in Man. This lack of adjustability may be the reason why such motor patterns, including ' consummatory acts,' are often so conservative in evolution that they can be used as taxonomic characters (4).

In the field of the ontogeny of behaviour,

therefore, many of Lehrman's criticisms are admittedly valid. However, if the ethological work done so far is stripped from its too sweeping generalisations and is put forward in more precise language, a core of distinct contributions remains which may be of great value to psychology; to ignore them would be to throw out the baby with the bathwater.

A few words must be said about Lehrman's remark concerning alleged abuse of the concept of homology by ethologists. When we call the mysterious mechanism responsible for the selective response to sign stimuli an 'Innate Releasing Mechanism,' and apply this term to such diverse animals as insects, fish, and birds, we do no more than give a common name to systems that, so far, have been shown to work in roughly the same way. The common term is by no means intended to indicate homology; it does not even mean that in all cases exactly the same mechanisms are at work. It is a functional term, just as our application of the name 'eye' to light receptors in vertebrates and insects is a functional term, which is not meant to indicate homology. As has been repeatedly argued by Lorenz (4) and others, homologising is only done with motor patterns found in groups of very closely related species. Lehrman's criticism reveals a lack of appreciation of the distinction between analogy and homology as applied with such great success in comparative zoology.

The Qualitative Study of Motivation

The ethological approach has also led to a new type of study of motivation. A thorough descriptive knowledge of motor patterns allows the observer to recognise the type of motivation prevalent in an animal at any moment. An experimental reconnaissance of underlying causal factors allows us to predict, from a knowledge of such factors present at any moment, the type of movement to be expected. This approach has led to the unravelling of the motivational situation not only in such activities as feeding, mating, escape from predators, etc., but also in those commonly labelled as 'display' (6). This work has clarified problems of motivation in agonistic displays and in courtship. Here we have to do with the interaction of two, and often of three, different types of motivation—usually attack, escape, and sexual motivation. This interaction gives rise either to combinations of 'intention movements' in the form of a mosaic, or to 'displacement activities.' The interaction of these different qualities of motivation causes an internal conflict, and it seems to me that these studies of conflict in animals may help us in understanding the effects of conflicts in Man. From the work of Liddell (3) and others it seems that the conflict behaviour of animals has much in common with neurotic phenomena, both in animals and in Man.

References

(1) Grohmann, J. (1939): 'Modifikation oder Funktionsreifung?' *Z. Tierpsychol.* **2**, 132–144.

(2) Lehrman, D. (1953): 'A critique of Lorenz's theory of instinctive behaviour.' *Quart. Rev. Biol.* **28**, 337–363.

(3) Liddell, H. (1938): 'The experimental neurosis and the problem of mental disorder.' *Amer. J. Psychiatry*, **94**, 1035–41.

(4) Lorenz, K. (1941): 'Vergleichende Bewegungsstudien an Anatinen.' *J. Ornithol.* **89**, Sonderheft, 119–229.

(5) Tinbergen, N. (1951): 'The Study of Instinct.' Oxford.

(6) Tinbergen, N. (1952): '"Derived" activities; their causation, biological significance, origin, and emancipation during evolution.' *Quart. Rev. Biol.* **27**, 1–32.

(7) Tinbergen, N., and Perdeck, A. C. (1950): 'On the stimulus situation releasing the begging response in the newly hatched Herring Gull chick' (*Larus a. argentatus Pont.*). *Behaviour*, **3**, 1–38.

(8) Weiss, P. (1941): 'Self-differentiation of the basic patterns of co-ordination.' *Comp. Psychol. Monogr.* **17**, 1–96.

The Modifiability of Instinctive Behaviour

by

Dr. R. A. Hinde

The aim of this paper is to sketch briefly some of the ways in which behaviour, which has developed without the intervention of 'learning' processes, may subsequently be modified by them. Since there has been considerable controversy recently about the nature and relative roles of learning and other processes in the development of animal behaviour, two examples of the 'unlearnt' components of

behaviour, whose modifiability we shall consider, will be cited first.

The first concerns the 'mobbing' response given by chaffinches to owls. When chaffinches, reared under conditions such that they had never seen an owl, were shown a stuffed tawny owl when about twelve weeks old they gave a characteristic, fully-formed mobbing response. This response is thus independent of any previous experience of owls. Further, such birds respond more strongly to a model which is roughly owl-shaped than to a similar but square one; and more strongly to one which has some indications of a face than to one which has not (Hinde 1954a). As a second example, canaries were reared in cages in which no nest material was present, and were not given opportunity to carry anything about in their cages other than seeds. In their first spring they went through the motor patterns of building a nest in a bare nest pan, even though they still had no nest material. When material was introduced, they came and picked it up within four seconds, and carried it to the nest pan within three-quarters of a minute. Subsequently they built tidy nests. (Hinde, unpublished experiments.)

Here we see that a particular type of response to a particular stimulus situation occurs without previous specific practice, conditioning or imitation. Such behaviour can conveniently be described as 'unlearnt'. The use of this term does not imply that learning (as that term is ordinarily understood) necessarily played no part in the development of elements of the behaviour (e.g. the development of general sensori-motor skills); nor, of course, does it deny the role of environment, acting through processes other than learning, on its ontogeny (*cf.* Lehrmann, 1953).

The effect of learning processes on a number of features of such a stimulus-response sequence will now be considered. The features chosen are: (*a*) the stimuli which elicit the behaviour in the first instance, (*b*) the stimuli which guide the behaviour in space, (*c*) the motor pattern, and (*d*) the consummatory stimuli to which the animal is particularly responsive while showing the behaviour: when these are perceived, the animal usually shows a reduced tendency to perform the response or behaviour in question. In the case of appetitive behaviour the consummatory stimuli are often also the eliciting stimuli for the next phase of the behaviour, whereas with consummatory behaviour they are the stimuli which we think of as causing a fall in 'motivation' (like food in stomach—see references in Hinde, 1954c). We shall also consider (*e*) changes in responsiveness to a constant stimulus.

(*a*) Changes in the stimuli which elicit a particular response have, of course, been studied in the context of classical conditioning, trial-and-error learning and so on. We are, however, only just beginning to gain any real understanding of the way in which such learning processes are integrated with processes which do not involve learning, to produce the adaptive behaviour seen in nature. As long ago as 1912 Wallace Craig showed that young doves, which had never drunk before, showed no tendency to drink at the sight or sound of water: only the presence of water inside the mouth elicited the reflex of swallowing. In these doves the first drink probably results from the bird 'accidentally' pecking at a drop or pool of water: subsequently the bird learns very rapidly the characteristics of the container or other surroundings of the water and, later still, presumably, the characteristics of water itself.

Thorpe and Jones (1937) have shown that the olfactory preferences, as indicated by the egg-laying behaviour, of an ichneumon fly can be influenced to some extent by the early (larval) environment. Adults have a strong genetically fixed tendency to approach the odour of meal-moth larvae, so that 85 per cent. of adults reared on meal-moth larvae turn towards the odour of meal-moth larvae when this is presented with the odour of wax-moth larvae in an olfactometer. Rearing on wax-moth larvae, however, causes a reduction in the number turning towards the meal-moth odour to 65·8 per cent. (The results of this study were cited incorrectly by Lehrman, 1953.)

Such studies as these indicate the importance which early experience may have in the development of behaviour which appears at first sight to be unlearnt. Many

other cases have been reviewed by Beach and Jaynes (1954).

One particularly dramatic class of modifications to the stimuli which elicit a response are provided by the phenomena known as imprinting. The classic case is provided by the grey lag gosling which will respond to the first relatively large moving object which it sees by following it and behaving towards it as though it were its parent. Later, sexual responses may be directed towards similar objects. (Heinroth, 1910, though see Spalding, 1873, for earlier detailed experiments.) These observations have been confirmed and extended to other species by Lorenz (1935), Fabricius (1951) and others. An entertaining case, described by Räber (1948), concerns a domesticated turkey cock. When adult, this bird would usually behave aggressively towards women, but would court men. Experiments showed that its behaviour was governed by an ' imprinted ' tendency to respond to human beings as members of its own species, but that humans were treated as male or female turkeys according to whether they did or did not carry any hanging or flapping articles about them. Usually such articles were carried by women, who were therefore treated as male turkeys, but men with brief cases were treated similarly. Räber suggests that this response to ' something hanging ' corresponded to an unlearnt response to the hanging wattles and drooping wings of a displaying male.

Although imprinting has been claimed as a ' special ' type of learning, it seems probable that it is not fundamentally different from other forms. Thus Lorenz (1935) claimed that imprinting was rapid, irreversible, and limited to a very definite and brief period of the life cycle ; but these are relative matters and other cases of learning show similar characteristics. Since goslings which have become imprinted on man will sometimes also court man when they become adult, it has also been claimed that imprinting is peculiar in that it may be completed long before the particular behaviour, to which the imprinted pattern will ultimately become linked, is established. No really adequate experiments seem to have been done here, however, for in all cases known to the writer the birds were more or less continually in the presence of man throughout their preadult life : the attachment of the sexual response to man, though perhaps influenced indirectly by the early experience, could have been due to a later learning process. In any case, a learning process through which animals become attached to certain objects as parents could hardly be *characterised* by the fact that sexual responses are later directed towards similar objects. That juvenile experience may influence patterns of behaviour which are as yet undeveloped is clearly shown by the experiments on olfactory conditioning in ichneumon flies by Thorpe and Jones (see above), but a comparable phenomenon has yet to be demonstrated in avian imprinting studies.

Experiments with moorhens, coots and chicks which are being carried out by Dr. W. H. Thorpe, Miss M. A. Vince and the writer seem to confirm this view that imprinting is not fundamentally different from other forms of learning. The following response of these species seems to be qualitatively similar to that shown by geese, yet the learning is certainly reversible and is certainly not limited to the first few hours or days of life.

(*b*) We may now consider those stimuli which guide the orientation of the behaviour. Unfortunately, space permits only one example, which is taken from the invertebrates. Tinbergen (summarised, 1951) has shown how the digger wasp, *Philanthus triangulum*, which returns a number of times with prey to its nest hole, learns the position of the nest in relation to certain landmarks during a ' locality study ' made on leaving the nest. In one experiment a ring of pine cones was placed round the nest while the wasp was inside. On leaving, the wasp flew around the hole for six seconds and then left. The ring of pine cones was then moved one foot from its original position. When the wasp returned she went to the ring of cones in its new position rather than to the real nest. This was repeated thirteen times, the ring being moved to a new position each time ; and the wasp found the nest only when the original arrangement was restored.

(*c*) Modifications to the motor pattern. ' Fixed Action Patterns ' or ' Erbkoordinationen ' seem to be extremely difficult to

modify. The movements used by animals in locomotion, preparing food, nest building, fighting, courting and so on seem to be modifiable only in intensity or in relation to the eliciting or orienting stimuli. Dr. Tinbergen has already quoted studies proving that flying in birds may occur without practice in the use of the wings during the period between hatching and fledging.

It would, however, be wise not to be too adamant in denying the modifiability of such patterns, for no real attempts have so far been made to modify them experimentally. Further, Skinner (e.g. 1948) has shown that, by careful training, it is possible to make animals perform extremely complicated and apparently stereotyped movements—a fact, of course, well known to circus trainers. It seems probable, however, that it is only in the higher mammals that this sort of learning becomes of great practical everyday importance.

There is, however, one group of motor patterns which are often subject to extensive modification as the result of experience, namely those involved in the production of vocal utterances. Thus, amongst birds, in some species the song seems to be extremely rigid and unmodifiable, whereas in others it is very largely learnt during the individual's life time. The bullfinch, for instance, can be taught to add whole new tunes to its repertoire. The chaffinch, which is being intensively studied by Dr. W. H. Thorpe, provides a very interesting intermediate condition. The song of this species normally consists of three phrases. There is a restricted inborn basis consisting of an ability to produce a song of about the normal length of a crescendo series of notes concluded by a single note of relatively high pitch. All further details are acquired by individual learning, but the inborn basis is sufficiently selective to ensure that the bird does not normally acquire notes from other species. There is evidence that some of the further details of the song can be learnt in the first few weeks of life, before the bird is able to sing itself, while the finer details are learnt during a period in its first spring, when it is first attempting to sing against other Chaffinches (Thorpe, 1954).

McDougall (1936) has pointed out that the actions of uttering sounds are the only ones which are perceived by the animal's own sense organs in the same way as if they had been uttered by another individual, as well as by the usual proprioceptive, etc., channels. The difficulty which animals have in imitating other types of motor pattern may be related to the fact that they cannot see themselves perform.

(*d*) So far, very little is known about the possibilities of modifying the consummatory stimuli—chiefly because of the difficulty of distinguishing between modifications to the consummatory stimuli and modifications to the eliciting stimuli of the next phase in the sequence of behaviour—and it is possible to do little but call attention to the problem. But it would seem, for instance, that 'imprinting' must involve something of this sort. Thus most nidifugous birds have a special type of appetitive behaviour which they show when they lose contact with their parents, whose proximity normally provides the consummatory stimulus situation. 'Imprinting' must involve modifications to these consummatory stimuli as well as to the eliciting and orienting stimuli of the following response. The mechanism may, of course, be the same in both cases.

It may be noted that the stimulus situations which act as reinforcing agents in learning situations (whatever the precise method of action of the reinforcing stimuli may be) are also stimuli which produce a decreased tendency to perform the immediately preceding behaviour. Thus the problem of modifications to consummatory stimuli may not be unrelated to that of modifications to reinforcing stimuli, i.e. to secondary reinforcement.

(*e*) Finally we come to the question of changes in responsiveness to a constant stimulus. If a constant stimulus is presented to an animal on a number of occasions, the strength of the response evoked will often gradually change. There is no doubt that the processes involved in these changes in responsiveness are both diverse and complex, and may occur at many different stages between the receptor and effector processes. It seems possible that some sort of preliminary classification of them in terms of (i) their degree of permanence and (ii) the extent of their

specificity to the stimulus situation, may be of help in their further study. For instance, if a stuffed predator is shown to an adult chaffinch, the 'mobbing' response given by the latter gradually wanes. It is likely that at least two processes are involved in this waning of responsiveness. One is subject to fairly rapid recovery and seems to be more or less independent of the particular stimulus which evoked the response, while the second produces a more or less permanent decrement in responsiveness and is specific to the particular stimulus (Hinde, 1954b). In addition, it seems likely that a third process is also involved—one which is subject to rapid recovery but is specific to the *stimulus*. Such a process has been demonstrated by Prechtl (1953) in his work on the begging response of young passerine birds. He found that, at a certain age when the response could be evoked by both auditory and visual stimuli, repeated evocation with one stimulus led to a waning of the response which was specific to that particular stimulus: the response could still be elicited by the use of the other. Of course these categories of 'stimulus specific,' 'relatively permanent,' etc., are only broad working ones: they will need to be modified and subdivided in the light of further research. Further, in the cases cited above there may also be additional processes giving a tendency towards an *increased* responsiveness with successive presentations. Nevertheless the characteristics of relative permanence and stimulus specificity do seem to be useful tools for the further investigation of these phenomena.

In this paper a few examples of the types of behaviour modification which one finds in animals have been mentioned. The types of learning referred to are by no means exhaustive. In particular, the processes involved in the development of sensori-motor skills, and the 'higher' types of learning (e.g. 'insight' learning) have been omitted. The aim has been merely to emphasise the importance and diversity of learning processes, and the absolute necessity of studying them in parallel with the other processes involved in the development of behaviour.

In conclusion, three particular aspects of behaviour modifications which are brought home rather more forcibly to the field ethologist than they are to the laboratory psychologist may be mentioned.

First, the nature as well as the complexity of the things an animal can learn often seem to be determined by its constitution. For instance, Dr. Tinbergen has demonstrated that herring gulls learn to recognise their own young in a few days but never learn to recognise their own eggs. This has obvious biological significance, for the young often stray from the nest, whereas the eggs do not.

In other cases not only the situations where learning can occur, but also what is learnt, seems to be characteristic of the species. Thus if the eggs of an incubating herring gull are moved a little distance from the nest, the bird will sit on the empty nest site, not on the eggs. The bird has an extremely strong tendency to learn the locality of the nest and the presence or absence of the eggs is relatively insignificant.

Secondly, while laboratory-oriented comparative psychologists usually carry out their experiments using food, water or escape from punishment as a reinforcing agent (considering such nebulous rewards as 'knowledge of results' to be effective only to sophisticated human subjects), the attention of the ethologist is frequently being drawn to cases of learning without any such conventional rewards—where, in fact, the perception of a particular situation seems to act as a reinforcing agent. For instance, from the experiments on the following response of young moorhens and coots mentioned above, it has become clear that the bird learns to follow what it *does* follow: improvement of performance does not depend on any 'reward' at the end of the run.

The third point to be emphasised is closely related to the last. The ethologist, seeing animals in their natural environment, cannot fail to be struck by the fact that they are *curious*, they investigate their environment. As yet we know very little about the nature of this curiosity, which remains as a misfit in any neat classification of behaviour types we try to make (though Berlyne, 1950, has attempted to accommodate it within the Hullian framework). We do not know even to what range of objects curiosity in any given animal will extend: Thorpe (1951) has

stressed that it is rarely, if ever, absolutely general.

Montgomery (1954) suggested that the 'Exploratory drive is a relatively independent primary drive aroused by novel external stimulation' and showed that a rat's tendency towards exploratory behaviour was reduced when other strong conflicting tendencies were present. Some observations made on the chaffinch support this view. This species shows 'curiosity' behaviour towards owl models which are relatively poor in owl-like characters: good models elicit a 'mobbing' response. Similar curiosity is shown by young chaffinches, in which the mobbing response is not yet fully developed, to good owl models. It thus seems that curiosity appears when either the external or internal factors are relatively low: if both are strong, mobbing occurs.

This mention of curiosity seems a fitting point at which to conclude, because there is little doubt that its analysis is one of the most urgent problems awaiting both ethologist and psychologist.

References

Beach, F. A., and Jaynes, J. (1954): 'Effects of early experience upon the behaviour of animals.' *Psych. Bull.* **51**, 239–263.

Berlyne, D. E. (1950): 'Novelty and curiosity as determinants of exploratory behaviour.' *Brit. J. Psych.* **41**, 68–80.

Craig, W. (1912): 'Observations on doves learning to drink.' *J. Anim. Behav.* **4**, 121–137.

Fabricius, E. (1951): 'Zur Ethologie junger Anatiden.' *Acta zool. Finn.* **68**.

Heinroth, O. (1910): 'Beiträge zur Biologie, namentlich Ethologie und Physiologie der Anatiden.' *Verkl.* **5**, *Int. Orn. Kong.* 589–702. Berlin.

Hinde, R. A. (1954a and b): 'Factors governing the changes in strength of a partially inborn response, as shown by the mobbing behaviour of the Chaffinch.' I and II. *Proc. roy. Soc.* B, **142**, 306–331, 331–358.

Hinde, R. A. (1954c): 'Changes in responsiveness to a constant stimulus.' *Brit. J. Anim. Behav.* **2**, 41–55.

Lehrman, D. S. (1953): 'A critique of Konrad Lorenz's theory of instinctive behaviour.' *Quart. Rev. Biol.* **28**, 337–363.

Lorenz, K. (1935): 'Der Kumpan in der Umwelt des Vogels.' *J. Ornithol.* **83**, 137–213, 289–413.

McDougall, W. (1936): 'An outline of psychology.' London.

Montgomery, K. C. (1954): 'The role of the exploratory drive in learning.' *J. comp. Phys. Psych.* **47**, 60–64.

Prechtl, H. F. R. (1953): 'Zur Physiologie der angeborenen auslösenden Mechanismen.' *Behaviour*, **5**, 32–50.

Räber, H. (1948): 'Analyse des Balzverhaltens eines domestizierten Truthahns.' *Behaviour*, **1**, 237–266.

Skinner, B. F. (1948): 'Superstition in the pigeon.' *J. exp. Psych.* **38**, 168–172.

Spalding, D. A. (1873): 'Instinct, with original observations on young animals'. Reprinted (1954) in *Brit. J. Anim. Behav.* **2**, 1–11.

Thorpe, W. H. (1951): 'The learning abilities of birds.' *Ibis*, **93**, 1–52, 252–296.

Thorpe, W. H. (1954): 'The process of song-learning in the Chaffinch, as studied by means of the sound spectrograph.' *Nature*, **173**, 465.

Thorpe, W. H., and Jones, F. G. W. (1937): Olfactory conditioning in a parasitic insect: its relation to the problem of host selection.' *Proc. roy. Soc.* B, **124**, 56–81.

Tinbergen, N. (1951): *The Study of Instinct.* Oxford.

Ethology and Psychology

by

Prof. A. Rex Knight

My allotted task in this discussion is to consider some of the aspects and implications of ethology that seem of special interest and importance to psychologists, and, in carrying it out, I shall consider not only the specific contents of the two previous papers but also the general theoretical framework in which they are set. In the first part of my paper I shall refer to four matters about which ethologists and psychologists are agreed; in the second part I shall refer to three other matters about which there is considerable divergence of opinion.

I

The first of the four matters about which ethologists and psychologists are agreed is that both groups are interested in both animal and human behaviour. Fifty years ago psychology was often supposed to be exclusively concerned with human beings, and the suggestion is still sometimes heard that the only place for animal studies in psychology is as a preliminary means of trying out hypotheses about human behaviour which must finally be established by human studies. Similarly, since most ethologists are zoologists and carry out most of their work with the lower animals, some people suppose that ethology is exclusively concerned with animal behaviour, and that when Lorenz and Tinbergen and other ethologists refer to human behaviour they are straying outside their

field. In fact, however, each of these views is mistaken. Ethologists and psychologists are both interested in all behaviour—in behaviour wherever it occurs, or, as Russell (1952) has expressed it, ' in behaviour phenomena, wherever they are found.'

Secondly, ethologists and psychologists agree that the comparative study of behaviour, including the comparison of animal and human behaviour, is both possible and useful. Four years ago Tinbergen (1951) referred to ' the almost universal misconception that the causes of man's behaviour are qualitatively different from the causes of animal behaviour,' and no doubt this belief still has its adherents. But it is no longer as widely held as it was when Huxley and Wilberforce clashed at the Oxford Meeting of the British Association in 1860, and psychologists, no less than ethologists, may now be taken to accept the view that the comparative method is as appropriate in the study of behaviour as it is in anatomy, embryology, physiology or biochemistry.

Thirdly, in their search for the causes of animal and human actions, ethologists and psychologists both proceed on the principle expressed by Hinde (1954) in the statement that ' behaviour is mediated by the nervous system, and every particular pattern of behaviour is mediated by a particular nervous mechanism or pattern of nervous activity.' This principle is hardly ' axiomatic ' (although Hinde has so described it), because, after all, it has for centuries been the subject of vigorous, and even violent, philosophical debate, as, for example, when Huxley put it forward at the Belfast Meeting in 1874. But it is now generally accepted in the biological sciences, and is one of the presuppositions of current ethological and psychological work.

Fourthly, psychologists would not dissent from much of what Hinde has said in this discussion about the effect of learning on the stimuli that elicit an inborn response, about ' imprinting ', about the response-specific and stimulus-specific processes underlying the waning of responsiveness to a constant stimulus, or about the three aspects of behaviour that he stressed at the end of his paper. They would have no difficulty in sharing his view that, although, of course, much more work needs to be done on the modifiability of behaviour, the conclusions that he has reached on these questions seem to fit the evidence so far collected.

II

Although there is common ground between ethologists and psychologists in the matters I have mentioned, there are other aspects of ethology, especially of Lorenz's theoretical system, which some psychologists—and, indeed, some zoologists—find invalid or controversial. The three objections to which I shall particularly refer are among those that have recently been forcibly presented by Lehrman (1953) in a searching critique of ethological theory. They may be raised in the form of questions, and my first question is: *Is the ethologists' use of the comparative method sufficiently strict?*

The comparative method deals essentially with similarities and differences among phenomena that are alike in origin and fundamental structure. It is concerned with the comparison of what is truly homologous rather than merely analogous. Now, do Lorenz and Tinbergen and other ethologists confine their comparisons to behaviour that is homologous? Is not Lehrman right when he argues that sometimes they make the mistake of assuming or implying that functionally similar behaviour patterns must, for that very reason, be alike in origin and caused by fundamentally similar mechanisms?

Ethology involves *inter alia* four fundamental propositions: (*a*) there are inborn or instinctive movements, which are each based on the activity of an innate specific mechanism in the central nervous system; (*b*) the activity of this mechanism is normally inhibited by another specific mechanism; (*c*) in response to specific stimuli, (' releasers ' or ' releasing stimuli'), the second mechanism ceases to inhibit the first and thus acts as an ' innate releasing mechanism '; (*d*) besides the ' releasing ' stimuli, there are also ' directing ' stimuli, which activate still different mechanisms and direct or orientate the animal's behaviour.

In support of these hypotheses, and of the hypothesised mechanisms that they entail, Tinbergen (1951) compares functionally similar behaviour patterns in a

variety of species. Thus, to take only one example, when he deals with the difference between releasing and directing stimuli and between their underlying mechanisms, he compares the behaviour of *Daphnia* in gathering near the surface of water with a high carbon dioxide concentration with the behaviour of *Pieris* and *Vanessa* butterflies in alighting on coloured papers when essential oils are sprayed into the air. In the first case, he says, it is the carbon dioxide that releases the response and the light near the surface that directs it ; in the second case the releasing stimulus is the scent, and the directing stimuli are the colours of the papers. But, even if we agree that the behaviour of *Daphnia* and the behaviour of the butterflies are each to be analysed in terms of a fundamental distinction between releasing and directing stimuli, it is surely impossible for the underlying mechanisms in the two instances to be alike in origin and fundamental structure.

In general, it must be regarded as mistaken to suppose that behaviour patterns that are functionally similar are necessarily homologous, for there is plenty of evidence that, at different evolutionary levels, functionally similar behaviour patterns often derive from fundamentally different mechanisms. Lehrman (1953) gives the example of the amoeba and the newborn baby, both of whom move towards weak stimulation and away from strong stimulation—the amoeba as a whole, the child locally. In both instances the behaviour serves the biological function of bringing the organism into contact with food and out of contact with harmful stimuli. But, as Lehrman says, the mechanisms underlying the response in the two organisms ' manifestly must be very different ' ; thus, though the behaviour patterns may be analogous, they are not homologous.

The second question is : *Are the ethologists' criteria of innateness satisfactory ?*

The concept of innate behaviour is basic in ethology. Tinbergen has described his work as ' the study of instinct,' and he and other ethologists construct their whole theoretical system on a foundation of inborn behaviour patterns and the inborn mechanisms that these behaviour patterns are held to involve. They are constantly seeking for behaviour which arises independently of the animal's experience and environment, or which, in Tinbergen's words in the present discussion, ' appears before conditioning and other learning processes set in.' It is therefore important to examine the criteria that ethologists employ in deciding whether an action is innate ; for, if these criteria are faulty, the basis of their system is unsatisfactory.

Lorenz and Tinbergen and their associates ascribe innateness to all behaviour which, in Hinde's phrase, ' appears without previous specific practice,' and, in particular, to any behaviour pattern which appears in animals that have been raised in isolation from fellow members of their species and prevented from performing the particular behaviour pattern in question. This is their principal criterion, and they use it frequently. Thus Tinbergen uses it as a basis for ascribing innateness to the pecking behaviour of the herring gull chick in response to the parent's beak (or a sufficiently similar cardboard dummy) ; Lorenz and Tinbergen use it in connexion with the egg-rolling movement of the grey lag goose; and Hinde uses it in connexion with the ' mobbing ' response of the chaffinch to a real or stuffed tawny owl (which, he says, ' can conveniently be described as inborn,' because it arises in a chaffinch ' under conditions such that it has never seen an owl').

Lehrman (1953), however, has pointed out that an animal isolated from others and prevented from practising a particular pattern '*is not necessarily isolated from the effects of processes and events which contribute to the development of the particular behaviour pattern.*' The important question is not ' Is the animal isolated ? ', but '*From what* is the animal isolated ? ' (Lehrman's italics). In this connexion he has referred to the finding of Kuo (1932) that the pecking behaviour of a newly hatched domestic chick, which certainly appears in chicks reared in isolation and therefore satisfies the ethologists' criterion, is the outcome of learning processes in the embryo. He has also referred to important unpublished work of Riess. Investigating the observation by Wiesner and Sheard (1933) that female rats raised in isolation normally build nests when they become pregnant, Riess not only raised rats in isolation but also prevented them

from ever manipulating or carrying any objects. 'The floor of the living cage was of netting, so that faeces dropped down out of reach. All food was powdered so that the rats never carried food pellets. When mature, these rats were placed in regular breeding cages. They bred, but they did not build normal nests. . . . They scattered nesting material all over the floor of the cage.'

Clearly, these and similar findings show that some behaviour, which ethologists, using their own criteria, would be bound to label as innate, demonstrably depends on the animal's previous experience and environment. Although it appears in animals reared in isolation, it does not arise *before*, but only *after*, 'conditioning and other learning processes set in.'

The third question is : *Are the ethologists on the wrong track in looking for unitary, autonomously-developing behaviour patterns?*

Many psychologists, especially in the United States, would answer 'Yes' to this question. They would hold that, apart from using unsatisfactory criteria of innateness, ethologists are mistaken in regarding the development of behaviour as involving the maturation of innate, specific, highly stereotyped behaviour patterns and of underlying specific neuromotor mechanisms, which belong, in their complete form, to the hereditary constitution of the animal. This is one of Lehrman's chief criticisms of ethology, and his criticism would be supported not only by those psychologists who reject every form of instinct theory, but also by those who, while averse from innate specific actions and mechanisms, readily accept both the existence and the explanatory value of instincts in the sense of innate drives or innate determining tendencies.

In general, psychologists, for heuristic and other reasons, prefer an approach which regards the development of behaviour as a process in which at each stage new patterns of nervous activity, and consequently new patterns of behaviour, emerge from interaction within the organism and between the organism and its environment. Thus Lehrman (1953) has argued that to label a behaviour pattern as innate throws no light on its development ; that 'the problem of development is the problem of development of new structures and activity patterns from the resolution of the interaction of *existing* structures and patterns' ; and that, though the physical growth of various structures is important in the development of behaviour, we must not suppose that such development involves the maturation of specific behaviour patterns or of their underlying mechanisms. As he puts it, the chick's pecking does not grow as a pecking pattern ; nest building does not grow as a nest-building pattern ; the elements out of whose interaction each activity emerges are not originally a unitary pattern ; in each case they become related as a consequence of their interaction during successive stages in the development of the organism.

There are other aspects of ethology that might have been discussed : for example, its account of 'vacuum activities,' of 'displacement activities' and of the hierarchical organisation of behaviour. But I should like to end on a grateful rather than a controversial note, and to say that, in spite of our imperfect sympathy with ethological theory, British psychologists are glad that the systematic study of animal behaviour is being revived in Britain, and they appreciate the part played by ethologists in this revival. The comparative study of behaviour owes much to the pioneering work carried out by Lloyd Morgan when he was Professor of Psychology and Education at Bristol at the turn of the century. But from that time until recently there were few studies of animal behaviour by British psychologists, who, now that they have again become actively interested in comparative psychology, look forward to increasing interaction and co-operation with ethologists.

References

(1) Hinde, R. A. (1954) : 'Changes in Responsiveness to a Constant Stimulus.' *Brit. J. Anim. Behav.* **2**, 41–45.

(2) Kuo, Z. Y. (1932) : 'Ontogeny of Embryonic Behaviour in Aves.' I and II. *J. exp. Zool.* **61**, 395–430 ; **62**, 453–489 ; III and IV. *J. comp. Psychol.* **13**, 245–272 ; **14**, 109–122.

(3) Lehrman, D. S. (1953) : 'A Critique of Konrad Lorenz's Theory of Instinctive Behaviour.' *Quart. Rev. Biol.* **28**, 337–363.

(4) Russell, R. W. (1952) : *The Comparative Study of Behaviour.* London.

(5) Tinbergen, N. (1951) : *The Study of Instinct.* Oxford.

(6) Wiesner, B. P., and Sheard, N. M. (1933) : *Maternal Behaviour in the Rat.* London.

PRESENTATION TO Mr. D. N. LOWE, O.B.E.

Mr. D. N. Lowe, O.B.E., resigned his post as Secretary of the Association on March 31, 1954, on taking up appointment as Secretary of the Carnegie United Kingdom Trust. He was appointed Assistant Secretary in 1935, the year in which the Association met at Norwich, and in 1946, on his return from war-time service at the Ministry of Production, he succeeded Dr. O. J. R. Howarth, O.B.E., as Secretary of the Association.

The years following the war were difficult ones for the Association and the marked success of post-war meetings was a tribute to Mr. Lowe's enthusiastic work and administrative ability. At the Oxford Meeting, in 1954, the General Committee appointed him an Honorary Life Member of the Association in formal recognition of his valuable services and of his devotion to the interests of the Association over a period of nearly twenty years.

Many Members of the Association have since joined together in making a gift to Mr. Lowe as a token of their affection and esteem. The gift, in the form of a cheque, was sent to him with a letter of appreciation from the General Officers and Council of the Association.

THE DIVISION FOR THE SOCIAL AND INTERNATIONAL RELATIONS OF SCIENCE

The Division, under the Chairmanship of Mr. Ritchie Calder, has arranged a two-day conference on ' Land Use,' which will take place, by kind permission, at the University College of the South-West, at Exeter, on July 15 and 16. The Conference will attempt to assess the present position after ten years of intensive planning and the effectiveness of the measures adopted, and will consider the following subjects : The Allocation of Land : the Administrative Functions of Central and Local Government ; The Need of Land : Agriculture and Forestry ; The Preservation of Amenities in Relation to Land Use ; The Need of Land for Mineral Exploitation, and Land Restoration ; The Needs of the Armed Forces for Land, and The Need of Land for Urban and Industrial Development.

The principal speakers will be Prof. S. H. Beaver, Mr. Geoffrey Clark, Mr. P. A. Macrory, Sir Herbert Manzoni, Sir George Pepler, Prof. L. Dudley Stamp, Prof. J. Sykes, Mr. George Langley Taylor and Sir James Turner. Mr. Ritchie Calder will sum up.

A detailed programme, including a form of application for tickets, may be had on application to The Secretary, British Association, Burlington House, London, W.1.

WATER CONTROL FOR INCREASED CROP PRODUCTION

At the Oxford Meeting of the British Association, Section M (Agriculture) devoted the morning of Thursday, September 2, to a discussion of water control for increased crop production. Dr. H. L. Penman, Rothamsted Experimental Station, read a paper on *Estimation of Irrigation Need.* Mr. E. L. Hoare, National Institute of Agricultural Engineering, then dealt with *Irrigation Techniques.* Mr. J. R. Fell, of Darlington, gave an account of *Grassland Irrigation* and, finally, Mr. H. H. Nicholson, University of Cambridge, discussed *The Prevention of Water-logging in Soil.*

Dr. H. L. Penman has prepared an integrated summary, which appears below, of the four papers.

Introduction

The object of water control is to keep the moisture content of the soil between a lower limit at which the soil is too dry to maintain full plant growth, and an upper limit at which physical or biological trouble may result. This upper limit may be reached by over-watering in irrigation, or by inadequate drainage maintaining too high a water table. The main purpose of this note is to try to decide what is meant by ' too dry,' how the condition can be changed on a field scale, and to indicate the benefit in crop yields when water control is achieved.

Theoretical Basis

From a physical point of view the essential features of transpiration are that the crop takes liquid water from the soil and this water reaches the atmosphere as vapour ; subsequently this vapour is mixed with the air above the crop. There are thus two independent lines of physical attack on the problem. First, recognising that the change of state from liquid to vapour demands a supply of energy as heat of vaporisation, the problem is one of energy balance. If energy income can be estimated, and if all ways of energy expenditure other than as latent heat can be estimated, then, by difference, the energy used in transpiration can be estimated. As the latent heat of vaporisation is a precisely known physical constant, it is easy to convert the ' energy used ' into ' water transpired.' This is known as the ' energy balance ' or ' heat budget ' method. Second, because the vapour has to be carried away from the leaf surface it is possible to calculate the rate of transfer from simultaneous measurements of vapour pressure gradient and of wind velocity gradient above the crop, the latter determining the degree of turbulent mixing in the air. This is known as the ' aerodynamic ' method.

Rothamsted work has shown that an instrumental difficulty common to both methods can be overcome by a combination of the two methods ; and that an estimate of ' potential transpiration rate ' can be made with an adequate degree of accuracy from standard weather data. The quantities required are :

Duration of bright sunshine,
Mean air temperature,
Mean air humidity,
Wind speed.

The potential transpiration rate so computed is the amount of water transpired in unit time by a short green crop of about the same colour as grass, completely shading the ground, of uniform height, and not short of water. It is not affected to any significant extent by rate of growth, for there is evidence that for a given crop the transpiration is unaffected by manuring that produces a two-fold increase in yield ; nor is it markedly dependent on variety

of crop (apart from colour) for there is evidence to show that several different plant types exposed to the same weather have the same transpiration rate.

In practice it is found that the dominant weather factor in Great Britain is the amount of solar radiation reaching the crop.

Theoretical reasoning and experimental results combine to indicate that the potential transpiration rate is a weather-dictated quantity, and as it can readily be calculated from weather data exploitation is possible in several ways. Two of these concern irrigation. First, from past weather records it is possible to estimate, for any chosen site, how frequently summer rainfall has been inadequate to meet the total potential transpiration for the summer, and thence, after making allowance for stored water in the soil at the beginning of the summer, to estimate how frequently irrigation would have been needed to prevent a check to full transpiration. A survey of this kind has revealed a *meteorological* need for more summer water arising in five years out of ten in England, south-east of a line Hull-Bristol, increasing to nine years out of ten round the Thames estuary. Second, by using current weather data it is possible to control current irrigation operations in experiments designed to see how far this meteorological need is also a farming need.

Conditions in glass-houses are sufficiently different for an independent treatment to be desirable, though the outdoor estimate of transpiration is a useful first approximation to what happens under glass. At the National Institute of Agricultural Engineering, crops of lettuce, tomatoes and carnations have been grown in large boxes frequently weighed, and it has been found that the transpiration rate is very closely correlated with solar energy passing into the glasshouse. The production of a cheap, efficient solar radiometer would be invaluable in control of glasshouse watering.

Technique

The relation between potential transpiration and weather gives a way of estimating the total amount of water needed to meet the full demand of the growing plant, but to avoid waste the possibility of rain must be allowed for, and only experience (much of which has still to be acquired) can decide how nearly the full demand should be satisfied. For example, full watering of some crops may delay ripening; while with others, notably horticultural crops harvested before maturity, it may make earlier harvesting possible. Water control reduces to attempting to manage the soil so that there is a deficit between total requirement and actual water applied. In the seedling stage this deficit must be small—say equivalent to not more than half an inch of water—but as the plant develops, bigger deficits can be tolerated. What is tolerable cannot be predicted; it must be found by experiment. Whether the control is by overhead irrigation, or by raised water-table, the deficit should leave sufficient margin for summer rain, to prevent run-off or waterlogging.

Cambridge experience on Burwell Fen has given convincing demonstration of this need when water-tables are high. Here the control has been exercised by drains, ditches and pumping, and it is not sufficient to provide a field with only one deep main ditch in very good order. The ditches all round must be in good order too.

Elsewhere control could be by irrigation, and at the outset it must be recognised that there is a basic problem of supplying water to the areas where the irrigation need is greatest. These are the areas where the local water supplies are least, and are already overtaxed in dry summers. Though in some places public mains can supply the water, a large area of land can only be irrigated from streams, so that farmers will be faced with the problem of pumping water. In general, if the equipment is to be mobile a diesel pumping set is most satisfactory; but if the pump is stationary, an electric motor is better. As example, Mr. Fell changed from a 60 h.p. diesel engine, giving 275 g.p.m., to two fixed electrically driven pumps each with a capacity of 325 g.p.m. feeding into permanent underground mains. Such mains are costly, and they must be big enough to keep frictional losses in the pipes down to reasonable limits: otherwise a lot of power is wasted in driving water through the pipes. The more quickly water can be applied, the

greater the use that can be made of the equipment, but, as it happens, the decision about speed is not primarily dependent on pumping economics. It depends on soil properties. Heavy watering will break down soil structure, leading to capping and consolidation that greatly restrict the rate of entry of water.

Experiments at the National Institute of Agricultural Engineering have shown that heavy clays clog quickly whilst coarse sands show little difference with change in droplet size. For an intermediate soil it has been found that small drops of 1 mm. diameter have little effect on infiltration rate, but large drops, of 5 mm. diameter, reduce the infiltration rate to only 7 per cent. of its initial value. This is on bare soil. Once a dense crop cover is established—or a mulch is applied—heavy watering is possible without damage.

Drop size is important also in the distance thrown. Depending on the pressure at the nozzle, there is a relation between the size of the drop and the distance it is thrown, a typical result being that sizes above 2 mm. diameter are received at distances beyond 30 ft. For small seedling crops, therefore, the ejection point should not be more than 30 ft. away or the drop size needed will become dangerously large. Irrigation of such crops is best done by fixed or oscillating spray-lines at about 40 ft. apart, or small rotary ejectors at about 35 ft. apart along the supply mains, so giving an overlap. For grassland, large rotary rainers can be used to cover 2 to 3 acres from one point : Mr. Fell puts down about 1 in. of water in 2 hours over 2 acres.

Results

(1) *Controlled irrigation*

Two groups of experiments have been carried out, using spray-lines, based on weather data from a meteorological station set up on the experimental site. The first, on two commercial farms in 1948, 1949 and 1950, was on sugar beet. Irrigation in 1948 was rendered superfluous by heavy rain after the waterings, but in 1949 and 1950 the calculated irrigation gave good increases in yield. As 1949 was a dry year, marked benefit from irrigation was not unexpected, but 1950 was regarded as being adequately wet for sugar beet. Calculations indicated a meteorological need for more water, and crop yields confirmed that this was a farming need. A summary of the results follows :

Sugar Beet : Kesgrave, 1950

Irrigation (in.)	Sugar (cwt./acre)	
	N_1	N_2
0	52	55
2·0	63	61
4·0	66	64

N_1=0·4 cwt. N/acre. N_2=0·8 cwt. N/acre.

The second set of experiments started at Woburn in 1951, on four farm crops : a grass/clover ley down for three years, and a three-course rotation of early potatoes, barley and sugar beet. Results of the first three years have not been analysed in detail, but their main import is clear. Although none of the summers has been very dry, and all have been wet from the end of July onward, there have been marked benefits on grass and early potatoes in all three years, a little on sugar beet in two years, and a little on barley in one year. As an example, the 1952 results for grass and early potatoes follow :

Woburn, 1952

Irrigation (in.)	Grass (dry matter) (cwt./acre)	Irrigation (in.)	Potatoes (tons/acre)
0·0	55	0·0	6·1
5·1	86	2·7	10·3

(2) *Empirical irrigation*

Since 1947 Mr. Fell has been irrigating 75–90 acres of temporary leys on his 240-acre farm in South Durham. The source of water is a sewage effluent drain taking treated sewage to the River Skerne near its confluence with the River Tees. Conservation is either as dried grass or as grass silage. The general aim is to give each field 1 in. of water as soon as possible after each cut from the middle of May onward, with at least four cuts per season of about 15 cwt. per acre per cut. The average production since 1947 has been about 4 tons of dry matter per acre, with an average protein content of 16 per cent. There are no available figures for non-irrigated fields, but the net output from the farm has increased by at least 45 per cent. since changing over, and this increase

has come from less than half the area of the farm.

In the eight years 1947–1954, the irrigation was found to be essential in five, of doubtful value in two (1948 and 1951), and unnecessary in one (1954). Costs have worked out at about £2 per acre per inch of water, the average annual cost working out at no more than what many people spend on fertilising an acre of potatoes.

The dried grass is fed as cubes to the dairy cattle, as meal in the pig rations, or sold. Since 1949 no concentrates have been bought for the cows.

(3) *Controlled water table*

In Burwell fen a range of steady ground water levels has been established in an area of fairly deep peat soil, in order to farm it with six crops each year, growing in a rotation over six years. Sample plots were harvested each year over the whole range of ground water levels, and the results were examined against the difference between the surface level of each plot and the mean ground water level of an adjacent observation well.

Results revealed the following trends :

(*a*) High ground water levels (15–20 inches) are capable of giving benefit to such crops as potatoes, celery, kale and sugar beet *if* the soil is initially in a good state, free from weeds, well drained, and summer rainfall is below average.

(*b*) Any level above 24 inches leads to failure of these crops in summers of excessive rainfall : and if maintained every year leads to difficulties in cultivation, weed invasion, poor tilth and eventually an appearance of dereliction.

(*c*) Below 24 inches yields are generally good for kale, sugar beet and rye-grass.

Conclusion

It has long been recognised that an adequate water supply is an essential requirement for plant growth ; and that too much water can be harmful, though there is singularly little quantitative evidence of the improvement of crop yields by drainage. Now it is becoming clear that in England the normal summer rainfall is not sufficient to maintain this adequate water supply, and that supplementary irrigation offers a way of increasing food production to an extent that, at present, is possible in no other way.

Selected References

Pearl, R. T. (Editor) (1954). ' The calculation of irrigation need,' *Tech. Bull.* No. 4. Min. Ag. and Fish.

Penman, H. L. (1953). ' The physical bases of irrigation control,' *Proc. 13th Int. Hort. Congress*, London, 52.

Penman, H. L. (1952). ' Experiments on irrigation of sugar beet,' *J. agric. Sci.* **42**, 286–292.

Bean, A. G. M., and Wells, D. A. (1953). ' Soil capping by water drops,' *Nat. Inst. Ag. Eng.* Report No. 23.

Fell, J. R. (1951). ' The use of intensive grass and green crop farming,' *J. roy. Socs. Arts*, **99**, 935–947.

Nicholson, H. H., Alderman, G., and Firth, D. H. (1951). ' An experiment in the control of the ground water-level in a fen peat soil,' *J. agric. Sci.*, **41**, 149–162.

Nicholson, H. H., and Firth, D. H. (1953). ' The effect of ground water-level on the performance and yield of some common crops,' *J. Agric. Sci.*, **43**, 95–104.

MODERN THEORIES OF COSMOLOGY

AT the Meeting of the British Association in Oxford in 1954 Sub-section A* (Mathematics) devoted the morning of Tuesday, September 7, to a discussion of modern theories of Cosmology. Mr. H. Bondi, Trinity College, Cambridge (now at King's College, London), dealt with Theories of Cosmology and he was followed by Mr. T. Gold, Royal Observatory, Herstmonceux, who contributed a paper on Cosmology as a physical science. Dr. D. W. Sciama, Trinity College, Cambridge, then discussed Evolutionary processes in Cosmology.

The papers by Mr. Bondi and Dr. Sciama are printed below.

THEORIES OF COSMOLOGY

by

Mr. H. Bondi

COSMOLOGY is the name of the subject that deals with the large scale properties of the universe as a whole. At first sight it may seem surprising that it should be at all possible to deal with such a subject in a scientific manner. For the essence of science is the possibility of observational disproof and it might be argued that no observational knowledge of the entire universe can be at our disposal for a long time to come, if, indeed, ever. The main purpose of this article is to correct this attitude, to show that cosmology can be treated scientifically and to indicate the main lines of thought now current in the subject.

Apart from some valuable considerations due to Newton, the first serious step in scientific cosmology was taken by the Hamburg astronomer Olbers about 130 years ago. In many ways Olbers' argument is the basis of all modern cosmology and it will accordingly be presented fully.

Olbers contemplated the fact that one saw in the sky a few really bright stars, a larger number of medium bright ones, and great numbers of faint ones. He argued that presumably, in general, the brightest stars were close to us, explaining thereby both their appearance of brightness and their small number, since there is relatively little space close to us. Similarly, he supposed the medium bright stars to be somewhat further away, accounting both for their smaller luminosity and their greater number. Finally, the faint stars would be really far away and hence be faint and numerous. But what about yet more distant regions of space? Should one not expect them to appear to be populated with exceedingly numerous but individually extremely faint stars? Olbers thought that this was so and accordingly he expected these immensely distant regions to contribute a faint background brightness to the night sky. He expected stars in the depths of space to provide an almost uniform glow in the sky, individual stars being too faint to be visible but so numerous as to provide, in the aggregate, such a glow.

Olbers was not content with this thought, but proceeded to calculate the background brightness to be expected. In this he followed the established scientific procedure of working out the observable consequences of hypotheses so as to be able to check them observationally. Since the individual distant stars were supposed to be too faint to be seen, Olbers could not rely on astronomical knowledge of the distant regions of space but had to make assumptions about them, assumptions that would enable him to infer the background light of the night sky and so to check the correctness of his assumptions. The assumptions Olbers made concerning

the constitution of the distant parts of the universe were as follows:

(i) The distant regions are essentially like our astronomical neighbourhood. The average distance between stars and the average luminosity of each star are more or less the same throughout the universe. In other words, the universe is homogeneous when viewed on a large scale.

(ii) The general character of the universe is not only the same at all places but also at all times. (This assumption is needed since, owing to the finite velocity of light, we see distant regions not as what they are now, but what they were like a long time ago.) In other words the universe is unchanging in time when viewed on a large scale.

(iii) On an average the relative velocity of any two stars vanishes, so that there are no major systematic motions in the universe.

(iv) The laws of physics as derived from our terrestrial experience apply throughout the universe.

These assumptions seemed to Olbers, as they seem now, to be the first and most obvious ones to come to mind. They are sufficient to deduce the corresponding background brightness of the sky. The actual derivation requires a simple and brief mathematical argument.

Imagine spheres to be drawn with the earth at the centre and of radii a, $a + h$, $a + 2h$, $a + 3h$ and so on, where a is much greater than h and h is a very large distance. The choice of h and a is determined by the need for every one of the spherical shells bounded by two successive spheres to be so large that it contains large numbers of stars (or, more correctly in the light of present knowledge, numerous galaxies). The number should be so large that average properties of luminosity and spacings can be used without serious error. Since the thickness of each shell is the same, the volume of each shell will be proportional to r^2, where r is the radius of the shell. (It is irrelevant whether the inner or outer radius is used since the thickness h is far smaller than r.) Accordingly, the number of stars in each shell is proportional to r^2 and so is the total rate at which stars in each shell emit light, since the average spacing and luminosity of stars have been assumed to be the same everywhere at all times. How much of this light reaches us? All the stars in the shell of radius r are at approximately the same distance r from us. Accordingly, the intensity of light received from any one of these stars is its rate of sending out light, divided by $4\pi r^2$. Therefore the intensity of light received from all stars in a shell is the rate at which all stars in the shell emit light, divided by $4\pi r^2$. Since this rate is proportional to r^2, it follows that the light received from a shell is independent of its radius (and does not vanish). Each shell contributes the same finite amount of light here. Since shells can be added without limit, the total amount of light received from all shells is infinite.

This result is not, however, quite correctly derived. True infinities rarely if ever arise in physical arguments. The discussion omitted to account for the finite size of stars which implies that each star not only sends out light, but intercepts the light of the stars behind it. This consideration does indeed prevent the sum becoming infinite, but, since stars send out so much light from a relatively small surface, the sum remains large. It turns out to be equal to the intensity of light on the surface of an average star which is about 40,000 times as strong as sun light when the sun is in the zenith.

This important result also follows from the consideration that, in Olbers' system, in whichever direction one looks, one's line of sight will eventually intercept a stellar surface. We are, therefore, entirely surrounded by stellar surfaces and so are at the same temperature and hence in the same radiation field. Alternatively, one can argue that the space between the stars forms a thermodynamic system that, by assumptions (ii) and (iii), has had time to settle down to equilibrium and hence the temperature everywhere in it is the same as that of the boundary, the surfaces of the stars.

The result derived from these assumptions is patently wrong. The earth is not immersed in a diffuse bath of light 40,000 times as strong as sunlight. Olbers' paradox, as it may be called, is in striking disagreement with observation. The derivation of the paradox is logically sound and

so the assumptions cannot be valid.[1] This is a remarkable result. The assumptions made concerning the nature of the universe have been disproved by observation The possibility of observational disproof is the decisive characteristic of science. Olbers' paradox, in demonstrating the possibility of the observational disproof of a set of cosmological assumption, shows cosmology to be a science.

But the utility of Olbers' paradox does not end with this achievement. The set of four assumptions cannot be valid, but might some of them at least be maintained? Which of them can be spared most easily ? There is no point in dropping assumption (iv) concerning the validity of the physical laws, since to throw away all our knowledge would seem to offer little hope of progress. Assumption (i) (Uniformity) has considerable direct and indirect observational evidence in its favour, so that the two assumptions remaining must be considered first. Can Olbers' paradox be resolved by dropping either of these ?

It is clear that this can be achieved by dropping assumption (ii). For if the universe is not unchanging, then one can postulate that the stars began to shine only a finite and not too long time ago. In this case we would not receive any light from distant shells since the stars there would not have been radiating at the time light would have had to leave them to get here now. This way of resolving the paradox may be stated briefly in the phrase 'The universe is young.'

[1] The only serious attempt at getting round the argument was made by Olbers himself who suggested that absorbing matter in space was responsible for our dark night sky. This suggestion is, however, incorrect, since absorption would raise the temperature of the absorber until eventually the absorber attained a temperature so high that it would radiate as much as absorb and so would be useless. By assumption (ii) it has had time to reach this state. Nor is it possible to resolve Olbers' paradox by supposing the geometry of space to be non-Euclidean. For all that is implied by this mathematical term is that the surface of a sphere of radium r is not $4\pi r^2$, but some other function of r. As Olbers' result only depends on the ratio of the surfaces of two spheres of equal radius being unity, the change is irrelevant. Similarly, even if the universe were finite, light could still encircle it arbitrarily often. The night sky would be illuminated by the indefinitely repeated images of a finite number of stars and so would be just as bright as in the Euclidean case.

The paradox can also be resolved if assumption (ii) is retained, but assumption (iii) is dropped. This is not quite so easy to see, though.

First, one must examine what motions exist that are compatible with assumptions (i) and (ii). It is not obvious that there are such motions, and some serious mathematics is required to find the answer. It turns out that the only motions compatible with assumption (i), *i.e.* the only motions of a homogeneous system preserving its homogeneity, are[2] those in which the relative motion of any two particles is along the line joining them, has a velocity proportional to the distance between the particles (if the velocity so derived is so large as to be comparable with the velocity of light, a more complicated formula has to be used) and is either inward or outward throughout. In the first case the entire system is contracting, in the second it is expanding. The rate of expansion (or contraction) is the ratio of velocity to distance, which is the same for all particles. If assumption (ii) (unchanging character) is made this rate must be constant, otherwise it may vary in time.

Having now found the motions that become possible by rejecting only Olbers' assumption (iii), we have to see whether either of them can resolve Olbers' paradox. Fortunately, terrestrial physics readily supplies the answer. It is known that light from a receding source appears to be redder and weaker than if the source were static, that these effects become large as the velocity of the source approaches the velocity of light, and that light from an approaching source is similarly bluer and stronger. If the universe is an expanding system, then the stars of the distant shells (in Olbers' argument) are receding from us, and the far distant shells are receding very fast. The light from these shells will therefore be considerably weaker than according to Olbers' argument, and the intensity of the background light will be less than calculated before. Accordingly, if the rate of expansion is sufficiently high the background light of the sky will be as faint as it actually is, and so Olbers'

[2] The additional, probably not very important, assumption of isotropy has also been made to simplify the result.

paradox will be resolved. If the universe were contracting then the light from distant shells would be enhanced and the paradox would be made worse.

It follows then that the darkness of the night sky, together with assumptions (i) and (iv), implies that the universe is either expanding or it is young or both these statements apply.

It is probably advantageous to discuss the modern observations at this stage before describing the current theories. A big telescope shows that the stars we see form an organised group, our galaxy. This is a disc-shaped object with a radius of about 40,000 light years (the average distance between stars is a few light years, which is a unit of distance equal to six million million miles) and a thickness of a few thousand light years. Looking out in the plane of the disc we see vast numbers of stars forming the phenomenon of the Milky Way. There are perhaps 100,000 million stars in our galaxy.

Looking further into space one sees, well beyond the confines of our galaxy, other similar groups of stars, other galaxies more or less similar to our own. These galaxies frequently occur in clusters containing anything from a few to a few thousand member galaxies. The average distance between galaxies (outside clusters) is towards a million light years, about thirty times the radius of an average galaxy. With modern telescopes galaxies can be observed even at distances of many hundreds of millions of light years, and hence vast numbers of galaxies are known. Allowing for clusters, the distribution appears to be reasonably uniform and so lends support to Olbers' assumption (i). A very difficult examination of the light of distant galaxies shows that the spectral lines are shifted to the red. The only plausible inference that can be drawn is that this reddening is due to a velocity of recession. It turns out that the red shift varies between different galaxies so as to be proportional to their distances from us (as inferred by the faintness of the light received) so that the velocity of recession is proportional to the distance. It was pointed out before that expansion with such a velocity-distance law is (apart from contraction) the only possible type of motion compatible with the assumption of uniformity. The fact that the observed motion is just of this type is a strong indication that assumption (i) (uniformity) is correct. Incidentally, the velocities measured by the red shift are very large (up to one-fifth of the velocity of light).

This is the setting in which the principal current theories of cosmology must be appreciated. The first of these gives up both assumptions (ii) (unchanging aspect of the universe) and (iii) (no motion), but bases itself firmly on assumption (i) (uniformity) and (iv) (validity of the terrestrially discovered laws of nature). Since the best formulation for the behaviour of matter-in-the-large is the general theory of relativity, this cosmological theory is based on it and is known as relativistic cosmology. General relativity and uniformity do not define a unique model of the universe, but a whole range of them. However, one of these, the model discovered by G. Lemaître, is generally thought to agree best with our universe. This is an evolving, changing model. The model starts off with all the matter of the universe in a highly condensed hot nuclear state, Lemaître's ' primeval atom.' The elements are made from the primitive hydrogen in a kind of nuclear explosion that leads to violent expansion. Under the influence of gravitation the rate of expansion slows down until an almost static state is reached in which the primeval uniform distribution of matter condenses into galaxies. A repulsive force that arises naturally in relativistic cosmology finally asserts itself sufficiently to start off an accelerating period of expansion, in which we now live.

The other current theory bases itself firmly on Olbers' assumptions (i) and (ii). Because of the stress it lays on the unchanging aspect of the universe it is called the steady-state theory. The reason for this stress is found in a critical examination of assumption (iv). The laws of nature, on which relativistic cosmology is based, are a summary of our terrestrial experiences gained in what is, on the cosmical scale, an extremely small region of space and an extremely brief period of time. It is impossible, on the basis of such limited experience, to judge which features are permanent and which are only temporary, or of only local significance. In particular,

if the universe varied in space or time it would be likely that some features of our physical laws (especially the so-called constants of nature) should vary with it. To put it differently, in such a varying universe there is no reason to expect our experiences to be typical, no reason to expect our summaries of these experiences, our laws of physics, to apply elsewhere. In these circumstances cosmology would be an exceptionally difficult subject in which every conceivable variation of the 'laws of nature' should be allowed for. The only possibility of avoiding these difficulties would arise if the universe were uniform in space and time, *i.e.* if Olbers' first two assumptions were valid. Then our experiences would not refer to any special place and time, but to typical conditions and hence would be applicable everywhere at all times. The assumption that this is so (*i.e.* (i) and (ii) valid) is known as the *perfect cosmological principle* as opposed to the so-called *cosmological principle* which states merely that the universe is, on the large scale, uniform in space, though not necessarily in time. It is the purpose of the steady-state theory to draw inferences from the perfect cosmological principle, inferences that can be checked observationally, so confirming or disproving the principle.

In the discussion of Olbers' paradox it was shown that if (i) and (ii) are retained, then the resolution of the paradox requires the universe to be expanding. In other words, since in the steady-state theory the aspect of the universe is unchanging, the resolution of the paradox given by the statement 'the universe is young' is inadmissible, and so the expansion follows. On the basis of the perfect cosmological principle the recession of the galaxies follows from the observation that it is dark at night, and the results of intricate astronomical work confirm this implication.

The next point is a little more difficult. If the universe is unchanging in the large, then all its large-scale physical characteristics (such as the density of matter) must be constant. But if the galaxies are receding then the intergalactic distances are increasing, matter is moving away from matter and so it would appear, by the law of conservation of matter, that the average density must be diminishing. This is in clear-cut contradiction to the perfect cosmological principle. The only way out of the contradiction is to suppose the law of conservation of matter to be incorrect to the extent that matter is being continually created. Although this may sound outrageous and contrary to experience, it is necessary to ask what the rate of this creation process is before ruling out this possibility. Now the mean density of matter in the universe is so low and the time scale of the expansion so large, that on calculation the mean rate of creation required to compensate for the effects of the expansion turns out to be one atom of hydrogen (or an equivalent mass) per quart volume every few thousand million years. This rate is obviously too small by many orders of magnitude to conflict with the experience on which the law of conservation of matter is based. There is therefore no need whatever to reject the idea of continual creation on the grounds that it disagrees with known evidence. It only disagrees with a mathematical extrapolation from evidence. Changing this may be inconvenient, but there is no reason to suppose the world to be arranged to suit the convenience of mathematicians.

In what form is this new matter created? This question is closely concerned with the problem of evolution. Every closed system is known to go through irreversible changes. Cosmically the most important of these is probably the conversion of hydrogen into helium, which takes place in every star, the excess energy being radiated away into space. Each system, each galaxy, is therefore ageing. How can the overall aspect of the universe remain unchanging, if every galaxy is evolving irreversibly? Only if new galaxies are being born, and old ones drift out of the range of telescopes through the expansion of the universe. The newly created matter must therefore stand at the beginning of the evolutionary chain, and, according to current astrophysics, this is cold diffuse hydrogen. The creation process must therefore imply the (presumably random) creation of hydrogen atoms of low velocity at a uniform rate.

The universe of the steady-state theory may be compared with a stationary

human population. Each individual is born, grows up, ages and dies, but the overall aspect of the population is unchanging. This requires not only that new individuals arise to take the place of the ones who have died, but that these new individuals should be babies, so as to stand at the beginning of the evolution of the individual. Similarly, the model of the universe is inhabited by galaxies drifting apart and ageing. In the growing spaces between the galaxies, continual creation leads to the existence of vast clouds of gas that condense into new galaxies which in turn age and drift away. All the large-scale average properties (density, distance between galaxies, size of galaxies, etc.) stay constant, though each member passes through irreversible changes.

This general description of the model of the steady-state theory leads up to the essential purpose of the theory, the discovery of crucial observations. There are several of these, but only a few can be discussed here.

(i) The light that reaches us now from distant galaxies left them a long time ago. If, as in Lemaître's model, all galaxies were formed at more or less the same time, then these distant galaxies must (on an average) have been younger when they sent out the light now received than near galaxies. On the steady-state theory, however, time does not matter. The average age of galaxies a long time ago was just the same as it is now.

The observations in question consist therefore in seeing whether there are any systematic changes in the characteristics of galaxies (intrinsic colour, shape, size, etc.) with distance. If there are any such changes, the steady-state theory will have been disproved; if there are none, it will be strengthened and relativistic cosmology will be weakened.

It is not easy to separate the effects of the red shift and of the faintness from changes of intrinsic characteristics. This difficulty does not allow definite conclusions to be drawn from existing observations. There is good reason to believe that further observational work will lead to a conclusive answer in a few years.

(ii) In the example of the human population it is clear that if the population is to remain stationary, the age distribution must follow a definite pattern. Correspondingly, the age distribution of galaxies must follow a certain pattern (with young galaxies predominating) if the steady-state theory applies. In Lemaître's model, on the other hand, the range of ages is quite limited, with no galaxy less than a certain age. Once it becomes possible to judge the age of a galaxy from its appearance, this test should be powerful. This possibility should arise before long.

(iii) and (iv) Tests relating to the origin of galaxies and of heavy elements. These will be discussed in D. W. Sciama's article.

Whatever may be thought of current theories, it is clear that they are not idle speculations, but serious attempts to discern crucial observational tests. The discovery of the cosmological significance of the darkness of the night sky made cosmology a science. Current theories, in asking for more intricate but entirely possible observations, are continuing this scientific path.

Evolutionary Processes in Cosmology

by

Dr. D. W. Sciama

A great complexity exists in the world. This complexity greatly complicates the physicist's task. Instead of matter consisting of a simple substance like hydrogen, it manifests itself in all the complex elements—about a hundred of them. And instead of being distributed uniformly it is clumped together into galaxies and systems of galaxies separated by distances large compared with their size. In many scientific problems such complications are a nuisance, but to the cosmologist they may provide valuable clues to the history and structure of the universe. The two examples I have given here have been the most analysed from this point of view. I want, therefore, to discuss how these complexities may have arisen.

The simplest possibility to state is that they came into being as such just as the rest of the universe did; there is nothing to explain. Such an answer is logically

possible, as is the suggestion that the whole universe has just been created, our memories and all. However, it is more fruitful to suppose that the complexities have developed from something essentially simpler. In particular, I want to explore the possibility that the complex elements originated as hydrogen, from which they developed as a result of nuclear reactions induced by high temperatures ; and that the galaxies have developed from some simpler distribution of matter owing to the action of gravitational forces.

This does not mean that I must restrict my considerations to a universe that was once nothing but hydrogen uniformly distributed, as is supposed in Lemaître's theory of an evolving universe. For example, in the steady-state theory of the universe, the newly created matter begins life as uniformly distributed hydrogen and evolves in a complex way because of the action on it of already existing complexities. Thus, in this theory the complexities are self-propagating. I should now like to discuss in a little more detail the processes envisaged by the two theories and I shall begin with the formation of the heavy elements.

A good deal of painstaking work has gone into building up our present picture of the abundance of the different elements in different parts of the universe. Studies have been made of the Earth, of meteorites, the Sun, the stars, interstellar gas and dust, and other galaxies. Objects which we cannot handle reveal themselves largely through their spectra ; the abundance of an element can be deduced from the strength of its characteristic contribution to the total spectrum. Unfortunately, this deduction is not a very certain one because one does not usually have a sufficiently detailed knowledge of the factors besides abundance which affect a particular element's contribution. Despite these difficulties, a fairly definite picture is beginning to emerge. Its most remarkable feature is that the abundance distribution is more or less the same everywhere; the surface of the Earth, the meteorites, the Sun, the stars, interstellar gas and distant galaxies all give very similar results when allowance is made for the special properties of each region. For instance, the surface of the Earth does not have its fair share of hydrogen, as gravity on the Earth is not strong enough to retain such a light gas.

The standard pattern is dominated by hydrogen which accounts for more than 95 per cent. of all the atoms in the universe —this is a further reason for starting everything off as hydrogen—the heavy elements which seem so important to us on the Earth are really only a slight impurity produced by special circumstances. Most of the remaining 5 per cent. is helium, such as can be formed from hydrogen in the interiors of normal stars. As we go to heavier and heavier elements the abundances drop very rapidly until we reach an atomic weight of about 100. After this the abundances stay about the same up to the heaviest elements. This regular general behaviour, and, indeed, the detailed small deviations from it, show that the abundances are related to the nuclear properties of the atoms and not to their chemical properties. Hence, it is in the field of nuclear reactions that we must seek the key to these patterns.

As I have already said, a high temperature is needed for nuclear reactions to take place ; only then do the particles collide with sufficient speed for transmutation to occur. Hydrogen gets converted into helium at the centre of the Sun and in similar stars, but elements heavier than helium need a temperature of at least fifty million degrees, which is more than three times the Sun's central temperature, and the whole range of elements need temperatures up to a thousand million degrees.

If now a critic insists that the stars are not hot enough to make heavy elements, I would reply with Eddington, 'Go and find me a hotter place.'

In the last few years Eddington's advice has been taken and two main candidates for the hotter place have been proposed : the universe itself in its first quarter of an hour, and certain special stars, particularly supernovae. The first possibility has led to what is known as the α-β-γ theory of element formation after its propounders Alpher and Gamow, who for obvious reasons implicated a well-known nuclear physicist called Bethe. This theory is set in a model of an evolving universe which started with an explosion and in its early stages had a high density and temperature.

Owing to its high temperature, the contents of the universe at this time were mainly radiation, the relatively small amount of matter present being in the form of neutrons. Radiation is not dominant at the present time because as the universe expands the density of radiation decreases faster than the density of matter. About a million years after the explosion the densities of the two were equal. Now after five thousand million years the density of radiation is only one millionth of the density of matter. Although the early neutrons have little influence on the universe in its extreme youth, their behaviour then determines the present abundances of heavy elements.

In order to determine how the neutrons behave, the α-β-γ theory uses the well-established results of nuclear physics. These show that the first step consists of many of the neutrons decaying into protons and electrons. These protons then absorb neutrons to form deuterons. The deuterons then absorb neutrons and so on. In this way heavy elements are built up, though it must be admitted that there is an obstacle to this process, which I shall mention later. When the nuclei become too neutron-rich, they adjust themselves by β-decay. Thus the build-up process consists of neutron capture followed from time to time by β-decay. In such a process the amount of any particular nucleus will be mainly determined by the ease with which it captures a further neutron, the more readily it does so the fewer of that type of nucleus will be left. In this way the theory provides a simple explanation of the observed distribution that I mentioned earlier in which the abundance falls rapidly as the atomic weight increases until a weight of about 100, after which the abundance remains constant. This is just what we should expect from experimental results on the ease with which different nuclei capture neutrons. After about fifteen minutes from the beginning of the explosion, the density of free neutrons has dropped so much that the whole build-up process comes to an end. So the observed abundance distribution now after five thousand million years corresponds to the state of affairs after only a quarter of an hour.

Those who advocate the steady-state universe have to look in special localities for their hot place, since the universe as a whole was never hotter than it is now. Indeed, they can only appeal to high-temperature processes that can now be observed.

The only regions known in which sufficiently high temperatures are generated are the supernovae—stars which explode with such violence that they often outshine the entire galaxy in which they are situated. The central temperature at the beginning of the explosion is believed to be about a thousand million degrees, which is of the required order of magnitude. The rate of occurrence of supernovae is roughly one per galaxy per century, so that as many as ten or a hundred million supernovae may have exploded in the Milky Way since it was formed. The exploding stars distribute their material throughout large regions of the galaxy, so that both the production and the distribution of the heavy elements can be accounted for. Furthermore, the explosion is so rapid that the abundances of the elements will correspond more to the initial high temperature than to the lower temperatures that occur subsequently.

A detailed study of the problem has recently been made by Hoyle, who has shown that a combination of the theory of stellar structure, of observations of stars, and of nuclear physics gives the right total amounts of heavy elements and the detailed way in which abundances vary with atomic weight. A striking success of this work was the prediction of a particular energy level of the C^{12} nucleus which was needed to obtain agreement with the observed abundances. Nuclear physicists at the California Institute of Technology were thereby provoked into seeking this level, which was, indeed, found at the predicted energy.

Now I want to discuss the way in which the two theories attempt to account for the other evolutionary process I have mentioned, namely the formation of galaxies. These objects seem to be the largest units from which the universe is built; each one contains a hundred million to ten thousand million stars. These galaxies are spread out through space separated by distances of about a hundred times their own size. This

distribution is itself far from regular, clusters of galaxies which appear to be gravitationally bound together being a prominent feature of the distribution. I have some slides which show what galaxies look like through the 200-inch telescope.

Now, how are these galaxies formed? The basic mechanism is presumably gravitation holding the material together or perhaps collecting it from an initially more widely distributed configuration. However, there are also opposing tendencies at work in the universe. Gas pressure is one, and another is the tendency for the whole system to expand. In order that galaxies should form, gravitation must overcome these expansive tendencies in localised regions, but not in the system as a whole. The two cosmological theories have characteristically different ways of achieving this.

The basic idea of the evolving universe theory is that there was one special interval of time when gravitation could win. The importance of gravitation relative to the expansive tendencies decreases as the density of matter decreases, so that at a later stage in the history of the universe galaxies could no longer form. This is the state we are in to-day. On the other hand, at a very early stage of the universe radiation was completely dominant and its dispersive effect on matter prevented the formation of galaxies. But it is suggested that there was a compromise time about a million years after the beginning of the universe when the density of radiation was sufficiently small and the density of matter sufficiently great. At this stage the gas is supposed to have broken up into blobs which remained the same size but which separated from one another, so that they are now much further apart. Since the present age of the Milky Way can be estimated independently, one can calculate how far apart the galaxies now ought to be, and one does indeed get an answer of the right order. The theory also gives a rough estimate for the mass of a galaxy in agreement with observation.

In the steady-state theory there is no moment of time with special properties ; galaxies must be forming at all times. The density of matter between galaxies is not too low for this because the creation of matter keeps it at a steady value despite the expansion. In this theory there is no first galaxy : any particular galaxy is born into a universe already full of galaxies. It is the galaxies existing at any time which enable all the dispersive forces to be overcome for the ones about to form. For the gravitational action of the galaxies on the gas around them perturbs it so as to produce concentrations which can then collapse under their own gravitational action. Thus these complexities are self-propagating. Obviously we require the children to have the same general properties as their parents, or the universe would not be in a steady state. This requirement that it must be an accurately self-propagating system is so stringent that it enables one to fix the properties of the galaxies entirely theoretically—their mass, their size, their distances apart, and their clustering tendencies with results in general agreement with observation.

I should like to end by making my own comparison between these two attempts to account for the composition and distribution of matter in the universe. I favour the steady-state theory for the following reasons. The α-β-γ theory of the heavy elements faces a severe difficulty. In the building up of the elements by neutron capture and subsequent β-decay there is a bottleneck : for there is no stable nucleus of weight 5, which means that the building up process stops at weight 4. The same difficulty would occur at weight 8. This difficulty was discovered by the proponents of the theory, who state that no way through this bottleneck has yet been found.

This difficulty does not exist in the supernova theory, where the density is sufficiently high for three-body collisions to be fairly frequent—these break through the bottleneck.

Furthermore, the steady-state theory only appeals to processes that can now be observed. These must in any case be taken into account, and if they suffice to explain the observed abundances any other substantial source of heavy elements is ruled out.

As regards galaxies, it has never been properly shown that they would form in the conditions envisaged by the α-β-γ theory, and owing to technical difficulties no detailed predictions have been made

about their properties and distribution. In the steady-state theory this can be done, and general agreement with observation is then obtained.

But to my mind there is a more important reason for preferring the steady-state theory. For in theories which start from an explosion the initial properties of the universe are entirely arbitrary. Thus it is possible to find an initial temperature which is favourable for making heavy elements, and then one simply has to assume that this was the initial temperature although the general theory would be equally compatible with any other initial temperature. This means that in this type of theory the laws of physics do not specify the contents of the universe, but only show how one state of the universe follows from another. The steady-state theory opens up the exciting possibility that the laws of physics may indeed determine the contents of the universe through the requirement that all features of the universe be self-propagating. This has already been done for the distribution of galaxies, and there seems to be no obstacle to extending this principle to other properties. The requirement of self-propagation is thus a powerful new principle with whose aid we see for the first time the possibility of answering the question why things are as they are without merely saying : it is because they were as they were.

THE NATURE OF THE FRICTION BETWEEN SOLIDS[1]

by

DR. F. P. BOWDEN, F.R.S.

THIS discussion deals with a very simple physical process : the sliding of one solid over another. The first thing we observe if we attempt to move one surface over another is, of course, that there is a resistance to the motion, which we call friction. If we maintain the motion for a considerable time we will also observe that the sliding surfaces are being worn away. These two associated phenomena exert a profound influence on our everyday life. Friction and wear, perhaps more than any other factor, limit the design and life of our machinery ; however, if the friction were greatly reduced we should find it extremely difficult to keep our cars on the road or even to walk. Our clothing is held together largely by the friction of its fibres : it is worn out, in many cases, by friction against other objects.

In order to understand the basis of the laws which govern the frictional behaviour of solids, the first step must be to examine the detailed nature of their surfaces. The more carefully one examines solids the more clearly one realises that their surfaces are never flat. We have used several delicate physical methods to examine the surfaces of solids and we find that no matter how carefully they are prepared they always have irregularities which are large compared with atomic dimensions. Recently, the new technique of reflection electron-microscopy has enabled us to obtain actual photographs at very high magnification of the fine details of the surface structure.

THE NATURE OF THE CONTACT BETWEEN SURFACES

Even with modern techniques of polishing it is difficult to prepare surfaces of appreciable size that are flat to within 100 or 1000 Å. Ordinary surfaces are very much more irregular than this. Thus whenever two solids are brought together contact first occurs at the points where the asperities on one meet the surface of the other. Initially the force driving the surfaces together is transmitted through these small areas of contact only, and consequently the local pressure is very great ; so great in fact that the summits of the irregularities deform plastically. They are crushed down until the area of contact is sufficiently large to support the load without the local pressure exceeding the yield pressure of the solids. Thus, although two metals may be apparently in contact over several square centimetres, in fact they will touch at only a few isolated points.

There is strong evidence that the friction between solids is due, in a large measure, to the adhesions which occur at the small areas of contact, and represents the force necessary to shear these junctions.

It is clear that this picture offers a simple explanation of the two classical laws of friction. First, since the real area of intimate contact is independent of the apparent size of the surfaces which are touching, the friction will be also. Secondly, as the frictional force is proportional to the true area of contact which is, in turn, directly proportional to the load, the coefficient of friction (which is the ratio of the frictional force to the normal force between the sliding bodies) will be a constant and will not depend on the load.

This local plastic deformation, which occurs when metals (and, indeed, most plastic solids) touch, causes the region of contact to become stronger than the bulk of the solid. Thus when the metals part

[1] A short summarised version of an Evening Discourse, delivered on September 6, 1954, at the meeting of the British Association in Oxford.

they separate not at the original interface but slightly to one side of it, in the region which has not been work-hardened; fragments become detached from one surface and remain welded to the other.

A convenient method of measuring quantitatively the amount of metal which is transferred at these localised regions of contact during sliding is to make one of the metals artificially radioactive and to measure radiation given off by the particles of it adherent to the surface of the other with a Geiger counter.

We have found that it is almost impossible to touch two metals together without causing some transfer from one to the other. The amount is greatly increased if any slip occurs, and also depends very much on the presence or absence of lubricating or other films on the surfaces.

Friction and Metallic Transfer in Surgical Operations

In simple mechanical operations such as screwing or bolting we may expect that some transfer will occur between the screwdriver and the screw or the spanner and the bolt. Although the transfer will be small it may, under certain conditions, be important. In modern surgery it is customary to use metallic plates as internal splints. These are screwed or bolted on to the bone and may remain in the body for long periods. It is known that if different metals are placed in contact in the human body they will corrode and that this can have a harmful effect on the healing of bone and soft tissue. Orthopaedic surgeons therefore take great care that all plates and screws are made of the same alloy. Nevertheless, it is found in practice that inflammation of the tissue near the buried metal does occasionally occur.

During an operation the orthopaedic appliances are handled and inserted with tools which are not made of the same stainless material. Now we have just seen that when two metals come into contact fragments of one may be transferred to the other. If this occurs in orthopaedic surgery it may be a possible source of trouble and danger. In collaboration with Mr. J. B. P. Williamson and Mr. P. Gowans Laing, we have investigated the amount and distribution of the transfer which occurs under these conditions using the radioisotope technique I have just described. Our experiments were carried out using samples of screws, plates, and tools manufactured for use by surgeons. We found that when a screw or bolt was inserted into bone and tightened using the procedure normal in orthopaedic surgery, appreciable amounts of the tools used to handle it became permanently embedded in its surface.

The metal used for the plates and screws is very resistant to corrosion, but that used for the screwdrivers and spanners is not. Measurements show that an appreciable difference in electrode potential can be set up between these two metals, so that the tiny transferred fragments could cause an electrolytic couple and may act as corrosion centres. This may result in a continuing dissolution of the buried metal which, although too small to be detected visibly, may be a cause of the adverse tissue reaction occasionally found surrounding orthopaedic components.

Recently analyses of samples of human tissue taken from regions adjacent to buried metal have shown that, even when there is no visible evidence of corrosion, the tissues adjacent to part of the metal handled with tools during insertion are abnormally rich in metal ions. Concentrations of chromium and nickel up to five times normal were found. The tissues near to unhandled parts of the buried metal, however, had concentrations not significantly above normal. This suggests that the fragments transferred from the orthopaedic tools during the operation may be a factor affecting the corrosion of the buried metal.

These physical observations may therefore help to explain the observed surgical behaviour and suggest that attention should not be focused on the buried metal alone. The transfer to it from the tool may be important and orthopaedic surgeons should give careful attention to the selection of their tools as well as of the inserted metal.

Friction and Wear in Space Ships

I should like to emphasise that the friction which we normally observe between metal surfaces (the force required to cause slip is usually between one-third and one times the normal load) is due

essentially to the presence on their surface of films of oxide or absorbed gas. These films are broken through to some extent under normal conditions of contact but nevertheless they play an important part in reducing the extent of the metallic contact and, in particular, of reducing the growth of the welded junctions when sliding starts. If metals are put in a high vacuum and these surface films taken off, it is impossible to slide them at all. The metallic junctions grow to a large size, the friction rises to a very high value, and gross seizure occurs. It is only when air is admitted that the friction falls to its usual value. In the almost complete vacuum of outer space unexpected frictional difficulties may be encountered. If Dan Dare wishes to move the joints of his space suit or pull the trigger of his Cosmic-Ray Gun he may well find they have seized solid, and that he floats helpless at the mercy of the more adequately lubricated Marsmen! Space engineers who wish to operate machinery on the surface of the moon, where there is little or no atmosphere, will need to take some pretty good lubricants with them.

Friction of Non-Metallic Solids and Development of Low-Friction Surfaces

There is evidence that the friction of non-metals is very similar to that between metals. Again localised adhesions occur at the small regions of actual contact and the friction is due mainly to the force necessary to shear these junctions. Usually the coefficient of friction is much the same as for metals, but I should like to mention briefly the behaviour of some special low-friction surfaces which may be of some practical interest and importance.

Molybdenum disulphide has a laminar plate-like structure (rather like that of graphite) which gives it a low friction. It is normally used in the form of a fine black powder welded on to a steel surface, but since it is only loosely attached it wears away easily. Instead of using a steel surface we have tried the experiment of using a molybdenum one and attacking it chemically so as to form the molybdenum disulphide *in situ*. It is then firmly attached and, if the molybdenum is porous, can penetrate to a considerable depth beneath the surface. Such surfaces can be heated; experiment shows that when running at red heat the coefficient of friction is still only 0·07 and is less than that of a well-oiled surface.

We have also studied the frictional behaviour of some of the new 'plastic' materials. One of the simplest of these is polytetrafluoroethylene (P.T.F.E.). This material has remarkable surface properties and exhibits a very low friction: unfortunately, it wears away rapidly. If P.T.F.E. is incorporated in a porous metal, such as sintered copper, we find that the resulting material has the mechanical and thermal properties of the metal and at the same time possesses the desirable frictional behaviour of the plastic. It maintains these properties up to high temperatures and is already proving to be a useful low-frictional material.

The Surface Temperature of Sliding Solids

The fact that intimate contact between solids occurs at only a few points has a second important consequence: the energy lost during sliding, which appears mainly in the form of heat, is all dissipated in these small regions. Theory suggests that these points may easily get very hot. We have used a variety of physical methods to measure this and find that it is indeed so. Although the mass of the solid appears to be cool, local high temperatures of 1000° C. or more occur very readily. These high temperatures can cause surface melting and play an important part in a number of physical processes: for example, in mechanical polishing, the surface of the solid is softened and smoothed by the heat generated at the points of contact with the abrasive particle.

Friction in Ski-ing and the Development of Some New Fast-Running Ski

Ice is an interesting substance; its friction can be lower by an order of magnitude than that of most other crystalline solids. Some years ago, in collaboration with Dr. T. P. Hughes, I made an experimental study of the friction of ice and snow at the Jungfraujoch Research Station. Our experiments supported the view expressed by Osborne Reynolds, about

1900, that the low friction is due to a surface melting. They showed, however, that pressure melting is a factor only at temperatures very close to the melting point of ice and with slow-moving surfaces. If the sliding speed is appreciable and at temperatures below 0° C. the film of water is produced, not by pressure melting, but by the frictional heating of the sliding surfaces. Under the conditions which obtain in ski-ing and sledging it is this frictional heating and melting which is the major factor responsible for the low friction. Experiments with sliders of different thermal conductance supported this view. Recently, I have extended this work to real ski-ing and I should like to mention briefly some of the results and, in particular, to describe the behaviour of some plastic materials sliding on ice and snow. The experiments supported the earlier conclusions that the low friction on snow and ice is due to a surface melting produced by the frictional heating of the sliding ski. At low temperatures the static friction of ice is high and it is not until the ski begin to slide at an appreciable speed that the friction falls to a low value.

The coefficient of friction of a number of different surfaces sliding on snow was measured. It was found that the friction for all the surfaces studied falls as the snow warms up and is least at 0° C. when surface melting can most readily occur. Very cold snow or ice behaves like any other crystalline solid and has a friction very similar to that of dry sand. The plastic P.T.F.E. gave a very low friction under all conditions investigated. I think this behaviour is due partly to its inherent good frictional properties and partly to the fact that it is not wetted by water. This suggests that P.T.F.E. suitably applied may have interesting possibilities for skiers or, at any rate, for those skiers who like to run fast. It is very easy to ski on as it gives a low uniform friction even when running on patchy snow. It may also have important applications to sledging and to aircraft landing, particularly on very cold, very wet or difficult snow or ice.

The Frictional Behaviour of Solid Krypton

Krypton is one of the inert gases and its solid state is of particular interest because of its inherent simplicity; it is held together by van der Waal forces only. In collaboration with Dr. G. W. Rowe we have studied its frictional behaviour in the temperature range from — 259° C. to its melting point at — 157° C. We find, as one would expect, that the solid is a very weak one. At liquid hydrogen temperatures it is about as hard as butter on a cool day. Nevertheless, at these temperatures, it gives a high fraction comparable with that of other crystalline solids. As the temperature approaches the melting point the sliding causes surface melting and the friction drops to very low values. The behaviour is very similar to that of ice.

Sliding Friction at 2000 m.p.h.

In present engineering practice speeds of sliding rarely exceed 50 m.p.h. What would happen if metals were slid together at speeds exceeding 1000 m.p.h.? Recently, in collaboration with Mr. E. Freitag, we have begun a study of the behaviour of metals and other solids which are sliding at these enormous speeds.

In order to do this we have made use of an elegant technique for suspending and spinning a sphere which has been developed for other purposes by Prof. Beams of the University of Virginia. A metal sphere is suspended by an electromagnet; it is illuminated so that its shadow falls on a photo-electric cell. If the ball sinks, the current in the magnet will increase; if it rises, the current will decrease. The ball can be rotated by applying a rotating magnetic field and, if it is enclosed in an evacuated chamber, can be made to spin at very high speeds (a million or more r.p.m.). The angular velocity can be measured accurately by observing the light reflected from an etch mark on the ball using a photomultiplier cell feeding into a cathode-ray oscilloscope. The surface of the ball can reach speeds up to 2000 m.p.h. and by bringing stationary surfaces into contact with it for a short time the frictional behaviour of metals at these high speeds can be studied. The results are striking. The coefficient of friction drops as the speed rises until, at 1200 m.p.h., the friction of copper on steel is one-third the value for ordinary

sliding. There is also a marked change in the nature of surface damage. As the speed increases the metals begin to show signs of surface melting and at extremely high speeds the frictional heat causes a melted layer which acts as a lubricant, in much the same way that ski melt a layer of water on ice. The wear per unit of distance slid can be very much less at these very high speeds than at normal speeds. Perhaps the technical problem involved in sliding metals at these enormous speeds may not be so grave. If only one could move fast enough one might perhaps ski on copper mountains !

A New Theory of Rolling Friction

Up till now I have considered the friction which arises when one surface slides over another. There is, however, a different way in which surfaces can move over one another : that is, of course, by rolling. It is common experience that it is much easier to roll surfaces along than to slide them ; the coefficient of rolling friction is perhaps less than 0·001.

The source of rolling friction has long been a matter of speculation. Osborne Reynolds showed that when a cylinder is rolled over a rubber surface it moves forward a distance less than its circumference for each revolution. He considered that minute slip occurred during the rolling and was responsible for the friction. More recently, Heathcote has discussed the slip which must occur when a ball is rolled in a groove.

Recently Dr. D. Tabor has made an experimental study of rolling friction and has suggested another mechanism which is quite different from these. He has demonstrated that the rolling resistance is almost unchanged by the presence of lubricants. This shows that slip is not an important factor. Tabor suggests that the primary source of rolling resistance arises from elastic hysteresis losses in the solids. As a ball rolls over a flat surface there is elastic compression of the region under it, followed by elastic recovery as the ball moves on. No solid is perfectly elastic and energy is dissipated in the hysteresis cycle. This represents the resistance to motion. He has demonstrated the correlation between the elastic hysteresis losses of materials and their rolling friction.

This work has shown clearly that rolling friction is not due to interfacial slip : it arises primarily from elastic-hysteresis losses. This accounts for the observation that the rolling friction is scarcely affected by the presence of lubricant films.

Conclusion

We have seen that the friction which is normally observed between sliding solids is due to the localised adhesion which occurs at the small regions of contact. With metals and with many solids complete adhesion is prevented only because the surfaces are contaminated by traces of adsorbed oxygen or other gases. If these films are removed strong adhesion occurs and the friction rises to an enormous value. It is, therefore, a curious accident that solids on the earth's surface have the friction to which we are accustomed. If it were not for this contamination much of our machinery would not move at all. Equally, of course, if the friction were much lower than it is the ordinary conduct of our lives would be impossible. One has only to step on a banana skin or drive a car on an icy road to realise this !

It is interesting to speculate on how the course of evolution would have been altered if the coefficient of friction had been a little different. The octopus for example has developed methods of moving over very slippery surfaces, and we might have had to do the same. On the other hand, had the coefficient of friction been much higher we might have had to lubricate our own path like a snail.

In engineering practice, of course, friction is a major problem. It is not so much the work we have to do in overcoming friction which is important, although this can be considerable; the real trouble is the damage that is done by friction : the wear or seizure of some vital part of the machine. Much ingenuity has been devoted to this problem and, up to a point, it has been successful : but only up to a point. One of the greatest limitations is temperature. It is difficult at present to run bearings hotter than 300° C. The whole tendency of our modern age is to

run things faster and faster and hotter and hotter; the engineer would like to run bearings and to slide surfaces at 1000° C. and at very high speeds. Why should he not? It means that new and unconventional surfaces and lubricants must be developed. A fuller and more detailed understanding of the physics of the sliding process under these extreme conditions will help to do this, and the problem, particularly with the advent of atomic power, is an urgent one.

ACKNOWLEDGMENTS

I thank members of my laboratory whose names appear in the text for permission to quote unpublished work.

A general discussion of the background of research in this subject is given in *The Friction and Lubrication of Solids* by F. P. Bowden and D. Tabor (Oxford University Press, Revised edition, 1954), and a fuller account of the work mentioned here is given in the Hawksley lecture of the Institute of Mechanical Engineers (1955).

LOCALISATION IN THE CEREBRAL CORTEX

AT the Oxford Meeting of the British Association Section I (Physiology) together with the Physiological Society devoted to the morning of Wednesday, September 8, 1954, to a discussion of *Localisation in the Cerebral Cortex*. The symposium was introduced by Dr. E. D. Adrian, O.M., P.R.S., President of the Association, and a past-President of the Physiology Section. Prof. E. G. T. Liddell, F.R.S., University of Oxford, read a paper on *Brain and Muscle* and he was followed by Prof. O. L. Zangwill, University of Cambridge, whose contribution dealt with *Speech and handedness*. Dr. F. P. Glees and Mr. J. Cole, University of Oxford, in a contribution which was illustrated by a film dealt with the effects of small lesions in the motor cortex of trained monkeys compared with similar lesions in the sensory cortex. Dr. J. A. V. Bates, Neurological Research Unit (National Hospital), examined the classical concept which associates voluntary movement with the cortico-spinal pathway in the light of some recent observations. The symposium ended with a contribution from Dr. E. A. Carmichael, Dr. M. R. Dix and Dr. C. S. Hallpike, Medical Research Council (National Hospital) on *Lesions of the cerebral hemispheres and their effects upon optokinetic and caloric nystagmus*.

LOCALISATION IN THE CEREBRAL CORTEX

by

Dr. E. D. Adrian, O.M., P.R.S.

THE question which is at the focus of our discussion this morning is why a large brain should be better than a small one for directing the sort of behaviour we call intelligent. It is true enough that ants and bees seem to lead a complex and successful social life with brains the size of pins' heads, but for the mammalian kind of behaviour a larger brain is needed and unless it is about as large as an apple we do not expect much evidence of our kind of reasoning and generalisation.

Naturally if we are looking for behaviour of the human type we should expect most sign of it in the animals whose brains are most like ours in their general plan. But absolute size seems to be an important factor as well as general arrangement and that is the basis of our problem.

One answer to it might be that a large size is needed to allow greater complexity of structure, that more different activities can go on in a large brain because there can be more subdivision into special departments, more localisation in fact. Another quite different kind of answer might be that intelligent behaviour demands a large bulk of undifferentiated nervous tissue to be at the service of whatever line of activity happens to be predominant.

Both answers are certainly correct up to a point and what we really have to decide is how much of the brain surface is permanently allocated to specific activities and what happens in the rest of it. And experience has shown that we can be very easily misled.

One hundred and fifty years ago it was widely believed that the cerebral cortex was divided into a number of regions each specially concerned with a particular faculty, with memory, cupidity, reverence, amorousness and so on. But the phrenologists were carried away by their theories and the tide set against them. Fifty years later, a hundred years ago, it was as widely believed that no distinction could be drawn between one part of the cerebrum and another. It was thought of, I suppose, as a large collection of cells like the liver, capable of many activities but equally capable in every part.

Then came Broca's demonstration of a particular kind of speech disturbance associated with localised damage of a small area of the cortex, and a few years

later there was Frisch and Hitzig's discovery that particular movements could be produced by electrical stimulation of particular cortical points. Localisation came back into favour. It was supported by the anatomists who found considerable differences in the connections and cell structure of different regions, and by the clinicians, who wanted some way of explaining the great diversity of symptoms which might follow cortical injury.

Fifty years ago the tide was running strongly in favour of a variety of specific centres in the cortex, each adding a distinctive element to the whole picture. There was the motor speech centre, the centre for understanding written words, the centre for appreciating spatial relations. For some of these there was clinical support but localisation was already too fashionable. The histologists were busy subdividing the brain surface into small areas of distinctive structure and the psychiatrists were finding correlations between the state of particular cell layers and the nature of the psychosis. They went too far and so now the tide seems to be ebbing again. We admit some localisation of structure and function in the cerebral cortex, but we are now very careful to guard ourselves from making too much of it.

The present position, I take it, is based on the view that the cerebrum is the supreme integrative organ, that its function is not to carry out a number of different and unrelated activities, but to direct the lines of behaviour of the organism to match the needs of its environment. Since it is at the highest controlling level it must bring all the relevant information to a single focus and the movements which it imposes on the muscles must have a single aim. One might almost say that a consequence of its function is that it cannot be split into departments. But the Cabinet which decides policy must have its Ministers, each dealing with a different field, and so in the organisation of the brain we can scarcely avoid finding some regions more closely related to one kind of activity than another. The information from the sense organs must come in by different channels and there must be separate channels outwards for the executive signals. We can trace these channels in the brain by a variety of methods, anatomical, electrical and clinical, and the most direct question we can ask is how far they go. Do they reach as far as the cortex and, if so, can they be traced further, to cross connections between one part of the cortical sheet and another?

The answer is 'Yes' to the first question and 'No' or 'Not yet' to the second. There is quite enough anatomical and electrical evidence to show that there are special channels up to the cortical level for the signals which bring information about different sorts of environmental change. The tactile organs on the body surface have a much more direct pathway to the post-central gyrus and to another small area near it than to the rest of the cortex; the visual signals go to the striate area and the auditory to the temporal and the precentral strip has more direct connections with the executive apparatus of the muscles and is more directly concerned with movement than any other part. There is no doubt about this sort of localisation at the cortical level. But although more or less definite pathways can be traced through the cortex, anatomically and by strychnine or electrical stimulation, we can only guess at their function. In fact our problems begin at the margins of what are called the projection areas, where the special channels come in and go out.

I think the problems have been made more difficult by our terminology. When I used to lecture every year on the cerebrum I had an eloquent passage on the muddle we were in from inadequate definitions—terms which could be used in different senses. I always gave some admirable definitions of my own. But when I examined the class in the summer I always found that most of them had either cut my lectures, or looked it up in the text-book, or had lost their faith to such an extent that they had given up the central nervous system as a bad job or remembered only some of the more misleading statements of the anatomists. Definitions may do harm by confusing our ideas, but we shall continue to misunderstand and misrepresent one another if we cannot agree on them, and terms like localisation and projection and representation have certainly been a fruitful

source of discord in the neurological camp.

An obvious example is that of the sensory area. The post-central strip is the arrival platform. It has special connections with the thalamus and in the anaesthetised brain we can easily detect the arrival of signals in it when different parts of the body are touched. These arrive according to a fixed anatomical pattern which maps or projects or represents the body surface in the post-central strip. But in Oxford, in 1916, Dusser de Barenne showed that a much larger part of the brain surface could give rise to sensory disturbance when it was stimulated by strychnine, and Penfield and Boldrey have shown the same thing in the human brain stimulated electrically.

For Dusser de Barenne, therefore, the sensory area extends widely over the cortex : its margins are indefinite though it is localised—it covers less than a third of the cortical surface ; but some of us still speak of the sensory area (or somaesthetic area) meaning the narrow post-central strip. The moral seems to be that we should use some term like the sensory receiving area when we are thinking of the region where the signals come in and another like the sensory arousal area when we are thinking of the area which can give sensation when it is stimulated. We cannot say that the function of translating sensory signals into sensation is localised to either. And it cannot be said that the function of the receiving area is altogether clear. There is the second sensory area with its own route from the periphery. It seems to be part of the organisation of the cortex, though we are aware of it only from its electrical reactions. And there is also the fact that the arrival of signals in the post-central area has been studied chiefly in the deeply anaesthetised brain. This shows the existence of a route which is not blocked by anaesthetics and it is certainly a permanent anatomical feature of the brain organisation, but in the active brain there may well be inhibitory influences which make the route sometimes open and sometimes closed.

In the other sensory fields, at all events in vision, we have the same kind of organisation, with a restricted area in which the signals arrive by the most direct pathway and surrounding regions which cannot be left out of the complete picture. In the motor field we know little or nothing about incoming signals and the evidence all comes from stimulation or injury. It is here, I think, that the confusion of terms is greatest, for instance when we argue whether movements or muscles are represented in the pre-central area and when we contrast voluntary and automatic movements.

But perhaps the disputes about the motor cortex give us the best example of the advantages which come from leaving definitions to look after themselves. Electrical stimulation of the brain is a valuable tool which has taught us a great deal about the motor cortex, but when we use it to find how the brain controls movement we must be very careful to keep our ideas from following too rigid a pattern. In spite of all its uncertainties the motor area was the first clear example of functional localisation : and it has remained the most fertile region for the localiser to investigate and the recent work from the Oxford laboratory shows how much it has still to teach us.

I can safely leave the motor field to Professor Liddell, but I ought not to leave out the fields which may well contain the answers to most of our questions about the brain, the association areas or the silent areas which have no connections except with other parts of the cortical sheet and give no significant response to electrical stimulation. These are the fields for the clinician, the neuro-surgeon and the electro-encephalographer. It is already clear that these large areas of the cortex are not all concerned with the same activities, but their effectiveness seems to depend more on their total bulk than on any details of their structure. They must be injured extensively for the brain to show any impairment and the impairment is often temporary as well as of a general rather than a specific quality. Perhaps we shall find that their function or part of it is also general rather than specific, a reinforcement of the prevailing activity by a large mass of cells which can add their momentum and join in the chorus.

But the Cabinet must meet to decide what should be done next and somewhere or somehow the different sorts of

information must be brought together and the appropriate movements must be set in train. We can say that the decision must take place at the highest level of the central nervous system, but we have not the slightest idea of its place in the cerebrum, whether its apparatus is concentrated or diffuse. Research on cortical localisation may lead us in the end to the supreme command, but at present I am afraid we are still entangled with the subordinate departments and it may be a long time before they will let us through.

Brain and Muscle

by

Prof. E. G. T. Liddell, F.R.S.

The problem of localisation of function in the brain has been of interest for a long time. Nearly a hundred years ago Hughlings Jackson, the great clinical philosopher of his time, deduced that the surface of the brain was 'irritable' and that irritation localised on the Rolandic area of the brain surface, by a spicule of bone or a new growth, produced the spasmodic movement of muscles which is called epilepsy—in this case 'Jacksonian epilepsy.' But as our President has indicated, it was Jackson's younger colleague, David Ferrier, who opened the new chapter of experimental studies by his well-known publication of 1873. It may be salutary for us to remember to-day that by 1890 William James, the psychologist, was already complaining about the numbers of published papers on the subject and the acerbity of discussion upon them. Jackson had observed, as I have already mentioned, that spicules of bone or tumours irritated the cortex. Ferrier in his experiments used electrical current for this purpose, as Frisch and Hitzig had done before him, though on dogs, while he went on to monkeys. Ferrier found that electric stimulation of localised parts of the cortical surface in the Rolandic region caused definite muscular responses in certain parts of the body and limbs, but not in other parts, though he did call attention to the fact that his 'areas' or 'circles' overlap, one with another, and from where they overlapped he might obtain a mixed response, but from the centre of the circles, only 'limited action.' This, at any rate, confirmed Hughlings Jackson's principle that the surface of the brain could be irritated or excited by external agency. Jackson said 'the artificial movements I have seen Ferrier produce simulate the movements of health.' Fifteen years later Horsley and Schäfer got broadly similar results and spoke of 'focal points' and 'circumjacent areas.' In one important respect Ferrier's results did not tally with Jackson's expectations because Jackson had for many years pointed out that in the type of epilepsy which is named after him the movements of the fit begin 'mostly' in the hand, foot and face. This often recorded observation implied that those parts of the body were widely represented on the cortex, but experimental electrisation of the cortex had failed to show that those parts had at all a particularly wide representation.

Looking back, as we can now do, upon this early work, it is not difficult to point out that results from different laboratories are likely to vary when different animals are used, when there are different anaesthetics and varying depths of anaesthesia, and different ways of applying electrodes to the brain. And in our observations of spinal reflex responses, which are elicited by repetitive electrical stimulation, there have often been recorded muscular contractions, as in the crossed extension reflex, with a long latent period, a slowly increasing tension and a rebound after cessation of stimulation. With natural stimuli, such as a pinch on the foot or pressure on the pad of the foot, there is a fairly quick response without rebound, but when the experimentalist stimulates a nerve with his unnatural electrical currents he is stimulating a mixed nerve which carries messages, some of which are excitatory and some of which are inhibitory, but the excitatory win for most of the time, and inhibitory messages show themselves in the long latent period, in the slow climb and in the rebound of the muscle's response. So, too, with repetitive electrical stimulation of the cerebral cortex. All too often the muscular response has then a long latent period, a slow climb and a post-stimulatory

rebound—features strongly suggestive of inhibition, masked by excitation—' latent inhibition ' as it has been called, due to simultaneous stimulation of excitatory and inhibitory elements. Furthermore when the experimentalist sets himself to stimulate the surface of the brain, and the cells in it, which lie close under the surface, he doesn't know what kinds of cells he is stimulating or how he is stimulating them. Is it reasonable to assume that only excitatory cells are stimulated ? It is well recognised that the surface area under which there are large cells is most easily stimulated to give responses in the muscles, but are these cell bodies, and all other cell bodies, e.g. inhibitory cells, stimulated directly or through by their sub-cortical dendrites ? There are a great many large cells which are not motor cells. What is their response to cortical stimulations ? In fact only a very few of the large cells are pyramidal motor cells. The work of Lassek is important here because he has shown that large pyramidal cells contribute less than 2 per cent. of fibres to the medullary pyramid. Then there is the possibility that satellite cells around the large cells may play a primary or a secondary part by repetitive response during electrical stimulation : and it is to be remembered that experimentalists have nearly always used repetitive stimuli (faradisation) which may easily create repetitive discharge among these constellations of cells. There is much we do not know. What the experimentalist at the very best is doing with his electrical blunderbuss is to break into the complicated machine of voluntary movement at a low level. His blunderbuss stimulates too unselectively and usually too widely. As his subject is generally an anaesthetised animal he cannot tell us anything about voluntary movement, but he can, nevertheless, often tell us of the reactions of that part of the machine which is internuncial between the cerebral level and the effector organ at spinal level, that is, muscle.

English experimentalists have worked mainly on motor phenomena, but as Fulton has pointed out they have not paid the same attention to the sensory function of the cortex as did their continental colleagues, of whom Munk was a leader in his time. As long ago as 1874 Hughlings Jackson said ' I do not shirk the logical consequences of the doctrine that the high centres are, like the low centres, sensori-motor,' and moreover with regard to the term ' centre ' he said ' it is an arbitrary expression.' Throughout his life he was to repeat these opinions over and over again.

Our President has referred to the experiments of Dusser de Barenne on sensation which were carried out in this laboratory some thirty years ago and in which little squares of filter paper, moistened with strychnine, were applied to the convexity of the cerebral hemisphere of *M. rhesus*. The limbs associated with that particular hemisphere became hyperaesthetic when the strychnine was applied to the sensory area, but just the same result was obtained when the strychnine was applied to the motor area. In fact, the motor area was fully sensory. I remember very well seeing these experiments in progress : how when Dusser de Barenne pulled gently on the hairs of the arm with dissecting forceps the monkey made no objection at all so long as the cortex associated with the stimulated limb had not been strychninised, but when the limb was approached with forceps after strychninising its cortex, the animal showed very great objection in anticipation of the stimulus. It was evident that that limb was hyperaesthetic, just as it was in the earlier experiments which he would have performed in Oxford, instead of Holland, had it not been for the intervention of the First World War, when the cats, which he was using then, showed by their behaviour that the affected limb was very hyperaesthetic. For some time these experiments stood alone as the most important contribution made by the experimentalist to the question of sensation in the motor cortex.

When it began to be possible to stimulate the brain of the conscious man without the use of an anaesthetic, as Penfield and Boldrey have done, it became clear that there was a wide area of representation in the cortex for the hand, face and foot. Their well-known work has made it quite clear that these parts of the body have such a wide area of representation on the cortex that there is very considerable

overlap of different areas. Moreover, sensation of different parts of the body has a very wide overlapping representation on the cortex in the same kind of way, overlapping also the motor area. The advantages of working with a conscious cooperative subject are, of course, considerably greater than those of working with a subject which is of lower order and under the influence of an anaesthetic. In recent years Dr. Phillips and I, here in Oxford, have repeated Ferrier's experiments of electrical stimulation in an anaesthetised animal but using a non-volatile anaesthetic which gives a fairly steady state of anaesthesia and does not give conditions of jerky anaesthesia such as is obtained with volatile anaesthetics. So long as our stimuli were repetitive, we obtained the classical results. But when our stimuli were single pulses applied once a second and the electrodes themselves were held in a mechanical holder, so as to narrow down error from vague manual applications, our results were different. We found in the first instance that with small currents the limited classical hand-area of the cortex when stimulated gives a movement which is a quick single flick of muscular action in the hand of very short latency, especially in the thumb and index finger, without rebound. When the strength of current is slightly increased this thumb-index area becomes much wider and about the same time that greater strength of stimulation will begin to elicit flick-movements from the big toe when the classical big toe area also is stimulated. Still stronger stimuli show that the thumb area increases very much in size so as to involve nearly the whole of the cortex. At the same time the area of the great toe is increased and these still stronger stimuli applied to the face area begin to give movement of the tongue and lips. Eventually there is obtained a thumb area so large that it overlaps the big toe area above, and the face area below, and these two latter areas themselves almost touch. The excitable area now covers not only the classical precentral gyrus but an area forward of it and an area behind it, on the post-central gyrus. The latency of action continues to be short and there is no rebound to suggest latent inhibition. Thus, for some reason which we do not well understand, there are these three parts of the body which as Jackson noticed in man, were most often involved in his form of epilepsy and must have a wide area of representation in the cortex in man, and the same state of affairs appears to obtain in the baboon, for that is the animal which we used. It is known also that these parts of the body in the cat have a wide representation in the cerebellum and, as most recent electrical observation of the cortex has shown, have also a wide sensory representation on the cortex. The important thing for us in the present time is to realise how many cells there are in the cortex about whose function we know so little. The number of cortical cells which perform the internuncial duty of conveying messages from the ' will-to-move ' to the spinal motor cells may be few indeed, but the co-ordination of their activities with sensory messages from their own muscles, from all other parts of the body and the sense organs, in short from the internal environment and from the external environment, occupies a vast number of units in the machinery of co-ordination. In fact, it is more than co-ordination. Cortical areas which are motor and cortical areas which are sensory are intermingled and overlaid one with another. It is Integration at its most magnificent : Integration which welds the diverse and diffuse nervous pathways of the cerebral cortex into one purposeful whole. (This intermingling and overlapping of allied tissue is now also in our picture of the respiratory centre, which is no longer the highly localised ' noeud vital ' of former days). We might well recall Hughlings Jackson's words on this kind of topic. He said ' it would be absurd to say that there is a buttoning centre. In the operation of buttoning there are slight and slow discharges of cells of different nervous arrangement. In consequence there is a harmony and a melody of the different movements of the hand and arm by which the button is got into its hole,' and while the button is being got into its hole all the other muscles of the body are alert and posturing and aiding the active muscles to perform their simple task. That is the Integrative Action of the Nervous System.

Speech and Handedness

by

Prof. O. L. Zangwill

The first step towards an understanding of the relation of speech to handedness was taken by Dax in 1836. This painstaking French neurologist was able to assemble no fewer than 140 cases of aphasia in which the lesion involved without exception the left cerebral hemisphere. But in view of the fact that Dax's monograph remained buried till 1865, when it was resurrected and published by his son, it had been customary to ascribe this discovery to Broca. As a matter of fact, Broca appears to have made it independently and was mainly concerned to delimit more precisely the area of destruction *within* the left hemisphere responsible for the loss of articulate speech. At a meeting of the British Association held at Norwich in 1868, Broca presented evidence that the area in question is the third frontal convolution—an area which has since come to bear his name. Although this evidence was far from convincing, being based on only two cases, Broca's correlation of so-called ' motor ' aphasia with lesions in the anterior region of the left hemisphere has since been abundantly confirmed and is generally taken as established.

It is altogether less well known that Broca was the first to appreciate the importance of *handedness* in relation to the cerebral localisation of speech. He clearly understood that whereas in the great majority of aphasias the hemisphere affected is the left, in a small minority of cases it is equally clearly the right. These relations were explained by Broca in the following way : In right-handed persons, he supposed that there is an innate functional predominance of the left hemisphere. In consequence, habits and skills acquired by experience, in particular those dependent on the higher intellectual functions, are built up, predominantly at least, under the control of this hemisphere. Now speech, as an acquired skill closely dependent on the intellectual functions, may likewise be expected to develop predominantly under left-sided cerebral control. In left-handed persons, on the other hand, Broca supposed that the position was exactly reversed : manual skill, and with it the acquisition of speech, are closely linked with the native pre-eminence of the right hemisphere. On Broca's view, therefore, handedness, and with it the unilateral representation of speech in the brain cortex, derives from an inborn predisposition of one hemisphere—usually the left—to take the lead in the acquisition of skill. This inferred functional pre-eminence of one hemisphere has come to be known as *cerebral dominance*.

As far as I can ascertain Broca did not himself put his hypothesis to the empirical test. The earliest observations which I have been able to trace of aphasia resulting from right-sided lesions in individuals whose handedness was reliably ascertained were made in England by Hughlings Jackson and other neurologists of his period. In 1870, Jackson was able to state that in nearly all cases of aphasia associated with left-sided paralysis which he had examined personally the patient has been left-handed ; and conversely, that right-sided paralysis in left-handed patients is not in general associated with aphasia. At the same time, he was careful to point out that not every patient with aphasia from a right-sided lesion is indubitably left-handed.

Broca's rule relating handedness and the laterality of cerebral lesions provoking aphasia appeared to be fully established by the end of the last century. Unfortunately, like so many rules in neurology, it was soon to be broken. Isolated exceptions multiplied with disconcerting rapidity. Bramwell, in 1889, devoted a whole paper to these exceptional cases, to which he gave the name of ' crossed ' aphasias. This term implies that the aphasia is produced by a lesion ipsilateral to the individual's preferred hand and, in Bramwell's experience at least, is not altogether rare among left-handed patients. Indeed he described a most convincing personal case of severe aphasia in a man of 35 who did everything save write with his left hand. The lesion in this case was almost certainly limited to the left hemisphere. ' Crossed ' aphasia in ostensibly right-handed patients has also been described from time to time in the literature, though not all these case-reports are wholly convincing. None the less, the work of Bramwell and other neurologists of his

period made it abundantly clear that Broca's principle allows of exception, particularly in the case of the left-handed.

Let us now turn to modern times. In 1935, Weisenberg and McBride reported a series of ninety-eight cases of aphasia. Of these, ninety-two were right-handed, two left-handed and three allegedly ambidextrous. Of the right-handed patients, all but one developed aphasia from a lesion in the left cerebral hemisphere. The exception was a woman with left-handedness in her pedigree and who was herself partially ambidextrous. Of the two left-handed patients, one developed an aphasia from a right-sided and one from a left-sided lesion. All three ambidextrous patients developed aphasia from lesions in the left hemisphere. Despite the small number of sinistrals in this series, there is some suggestion that aphasia is more likely to follow left-sided lesions in left-handed or ambidextrous patients than right-sided lesions in right-handed patients.

In 1936, Chesher summarised the evidence from 166 cases of circumscribed lesion within the speech areas of one or other cerebral hemisphere. Of these cases, ninety-four were left-sided and seventy-two right-sided lesions. Of the patients with left-sided lesions, all except one were aphasic—the exception being a left-hander. This group, however, included five predominantly left-handed patients who none the less became aphasic. Of the patients with right-sided lesions, all but six were purely right-handed and none of these became aphasic. Two consistently, and three predominantly, left-handed patients within this group also became aphasic, as did one patient who, although predominantly right-handed, appeared to favour his left for certain unimanual skills. Thus of the eleven naturally left-handed patients, five developed aphasia from a lesion ipsilateral to the preferred hand, i.e. were cases of ' crossed ' aphasia. On the other hand, there was only one case of ' crossed ' aphasia in a right-handed man and his hand-preferences, as we have said, were not unmixed. It would appear, then, that Broca's principle, while valid in general for right-handers, applies doubtfully, if at all, to those of sinistral bent.

A more recent study by Conrad of brain-injured German Service personnel in the last war gave very similar results. Conrad managed to collect forty-seven cases of brain injury in left-handers, of which eighteen showed some form of speech disturbance. The lesion was left-sided in ten cases, right-sided in seven and bilateral in one. Aphasia was severe and long-lasting in five cases with left-sided lesions and in an equal number of cases with right-sided lesions. In five left-sided and two right-sided cases the speech disturbance was transitory. Conrad's material also includes ten cases of lesion within the speech areas, equally divided between the left and right hemispheres, in which there was no aphasia. In short, we have five cases of aphasia following injury to the hemisphere ipsilateral to the preferred hand ; and five cases of injury to the hemisphere contralateral to the preferred hand which did not cause aphasia. This means that we have a further ten left-handers in which the hemisphere dominant for language is indisputably the left.

I would like to refer briefly to some similar studies we have recently made in this country. My former colleague Mr. Michael Humphrey, working in Oxford under the direction of Dr. Ritchie Russell, studied ten cases of unilateral brain wounds in left-handed Service men. His criteria of selection were stringent, only cases being included in which the lesion was strictly unilateral and so situated as to produce aphasia if, and only if, the affected hemisphere were the dominant one. Selection was thus in terms of the site of the lesion rather than the actual presence of aphasia. Of five cases with appropriate left-sided lesions, two displayed severe and lasting aphasia and three speech disturbances which, though initially severe, underwent a marked degree of restitution. Of five cases with appropriate right-sided lesions, aphasia was severe and persistent in two cases, transitory in one and absent in two. In at least four of the patients, therefore, the presumptive dominant hemisphere was again the left ; indeed seven of them may be said to have shown some tendency to ' left-brainedness ' in all their speech functions.

We have been carrying out some further

studies of left-handed patients with unilateral brain disease at the Psychological Department of the National Hospital, Queen Square. Nine patients have so far been examined and we hope to report the findings in detail at a later stage. For the present, it may be said that of seven sinistral patients with left-sided lesions, severe aphasia has been present in four, minimal aphasia in one and no speech disturbance in two. Of the two patients with right-sided lesions, one showed a severe, though rapidly improving, aphasia and the other no speech disturbance whatsoever. In this series, then, a majority of patients so far—five out of nine—gave evidence of left cerebral dominance in regard to speech.

TABLE I

	Number of cases	Left-sided Lesions			Right-sided Lesions		
		Severe Aphasia	Minimal or Transitory Aphasia	No Aphasia	Severe Aphasia	Minimal or Transitory Aphasia	No Aphasia
Weisenburg & McBride (1935)	2	1	0	0	1	0	0
Chesher (1936)	11	3	3	0	4	1	0
Conrad (1949)	27	5	5	5	5	2	5
Humphrey & Zangwill (1952)	10	2	3	0	2	1	2
Zangwill (1954)	9	4	2	1	1	1	0
	59	15	13	6	13	5	7

The general position regarding the laterality of the lesion causing aphasia in sinistrals is indicated in Table I. This is compiled from findings reported by Weisenburg and McBride, Chesher, Conrad and ourselves.[1] It will be seen that the number of cases in which there is left hemisphere dominance, as inferred from presence of aphasia with left hemisphere lesions or lack of aphasia with right hemisphere lesions, is a little larger than the number of cases in which right hemisphere dominance is similarly inferred. Thus the cases which conform to Broca's rule are actually outnumbered by the exceptions ! It will also be seen that about a third of the cases displayed an aphasia so mild or so transitory that no certain conclusions as to hemisphere dominance could be drawn.

The high frequency of ' crossed ' aphasia in sinistrals as compared with dextrals is difficult to explain at the present time. One suggestion, originally put forward by Bramwell, is that acquired factors may in rare cases override the influence of heredity in the determination of handedness. He pointed out that in most of the cases in which a left-handed person becomes aphasic as a result of a left-sided lesion his ancestors will be found to have been right-handed. This idea was later taken up enthusiastically by Foster Kennedy, who adduced what he called ' stock-brainedness ' as the causative factor in ' crossed ' aphasia. According to this author, a left-handed person occurring eccentrically in a right-handed stock is dominated in his cerebral organisation by the trend of the stock rather than by his own individual peculiarities. Conversely, a sinistral trend in the stock may produce, in a right-handed individual, a ' condition

[1] L. Roberts, in a recent survey of 563 cases operated upon by Penfield for the relief of epileptic seizures, reports that operations on the left hemisphere were followed by aphasia in the case of thirteen out of eighteen left-handed patients who had sustained their brain damage after the age of 2 years. On the other hand, only three out of twenty-three left-handed patients developed aphasia following operations on the right hemisphere. In view of the fact that Roberts gives no details as to the locus of the excisions or the type and severity of the aphasia, his data have not been included in the present report. It is, however, of interest to note that he considers the left hemisphere to have been dominant in practically all cases in this series, irrespective of handedness.

of brain ' similar to that of his collaterals. On this view, then, ' crossed ' aphasia is due to a dissociation between cerebral dominance—governed by genetic factors—and handedness—governed by some specific peculiarity of the individual.

The theory of ' stock-brainedness ' has found considerable favour among neurologists and a good deal of evidence can be marshalled in its support. Thus the rarity of ' crossed ' aphasia in dextrals can presumably be ascribed to the rarity of strongly left-handed stocks ; its frequency among sinistrals on the other hand, can be ascribed to the overwhelming predominance of right-handed stocks. None the less, this explanation is not without its difficulties. Thus it does not appear that a family history of left-handedness can *necessarily* be taken to mean that the right hemisphere is dominant in any given individual, even if he be himself left-handed. Cole has reported the case of a left-handed patient, whose mother and maternal grandmother were likewise left-handed, who developed severe aphasia from a lesion in the left cerebral hemisphere. Conversely, German and Fox reported the case of a left-handed man with sinistral pedigree who failed to develop aphasia following operative removal of the right frontal lobe. Similar cases have occurred in our own experience. Until the genetics of handedness and cerebral dominance are better understood, therefore, the concept of ' stock-brainedness ' must be accepted with a certain reserve.

A second explanation of ' crossed aphasia ' places greater weight on the influence of developmental and acquired factors in cerebral dominance, more especially enforced conversion to the use of the right hand in writing. Bramwell maintained that factors of this kind, especially when associated with a family tendency to left cerebral dominance, might well outweigh the effects upon the latter of an apparently native preference for the left hand. The left-handed neurologist Tilney presented a similar explanation of his own aphasia, which was particularly significant in so far as he came from left-handed stock but none the less developed aphasia following a lesion in his left hemisphere. In our own experience, it is certainly true that most of our patients with aphasia following left-sided lesions have been right-handed writers, but this is equally true of most of our cases in which the aphasia has followed a right-sided lesion. We have so far been able to collect only five cases of aphasia in left-handed patients who have always written with the left hand. Of these, two developed aphasia from left-sided and two from right-sided lesions ; the remaining patient developed but slight dysphasia from a left-sided lesion. It is at least plain, therefore, that language may be represented in the left hemisphere even in the case of a patient who (unlike Tilney) has always written with the left hand.

In considering the general relations of handedness to cerebral dominance, it must be borne in mind that the former is certainly not a specific genetic trait, comparable to eye-colour or skin-pattern. Although there is almost certainly some degree of inherited bias towards either right- or left-handedness, the actual handedness shown by an adult appears to be the outcome of a complex growth process in pre-natal and early post-natal life, modified perhaps by social training in later childhood and adolescence. As regards the earliest years, Gesell and his associates have been able to show that consistent hand preference develops rather slowly and is seldom fully established before the age of four years. Although signs of a hand-preference may show themselves early, phases of bilaterality and fluctuation commonly occur in the first three years of life. It would, therefore, appear that handedness, like the tonic neck reflex pattern out of which it evolves, is subject to many and varying influences in the course of early development. Indeed it is theoretically possible that factors of maturation and development may over-ride the influence of heredity in determining handedness.

It is equally clear that unilateral hemisphere dominance is far from fully established at birth. Broca hazarded the view that a lesion of the 3rd left frontal convolution acquired at birth or shortly after would not interfere with the normal development of speech. His brilliant guess has proved entirely correct. Injury or disease of the left hemisphere sustained

before the age of about two years neither impairs nor retards the acquisition of language. If, moreover, the affected hemisphere is later removed, the operation is not found to give rise to aphasia. In earliest infancy, then, there is a measure of equipotentiality of the two hemispheres with regard to cerebral dominance. Despite an inherited bias towards 'left brainedness,' language may in exceptional circumstances be acquired under control of the right cerebral hemisphere.

As I see it, the explanation of 'crossed' aphasia must run somewhat along the following lines: In the early growth of handedness, factors other than inherited bias play a real, though as a rule minor, role. In rare cases, the course of development may be supposed to favour the emergence of a left-handed preference despite an inherited tendency towards right-handedness. Yet in these cases, it would none the less appear that language development follows the normal rule, being governed, predominantly at least, by the left hemisphere. A dissociation between handedness and cerebral dominance must therefore result. If such an individual should later become aphasic, it will be from a lesion within the left cerebral hemisphere. In the same way, it may happen—if very rarely—that an individual of left-handed stock becomes right-handed through the vicissitudes of early development. Should such an individual later become aphasic, it will be from a lesion within the right cerebral hemisphere.

If my view is correct, at least half of the normal left-handed population are in reality 'shifted dextrals,' i.e. persons without family history of sinistrality who have acquired their left-handedness in the course of early development. Such an individual tends to show left cerebral dominance for language in accordance with his inherited bias, and it is at least possible that this dominance is strengthened if he be later taught to write with the right hand. In 'congenital' sinistrals, on the other hand, the hemisphere dominant for language is typically the right. In these cases, education of the right hand for skilled activities, more particularly writing, may lead to some measure of transfer of dominance to the left hemisphere. It is, however, most doubtful if this transfer is ever complete, except possibly in the case of writing itself. None the less, it is to be expected that many 'congenital' sinistrals, by virtue of early right-handed training, will come to acquire some measure of bilateral representation of their speech.

Can Voluntary Movement be Localised in the Cerebral Cortex?

by

Dr. J. A. V. Bates

For eighty years there has been a simple doctrine relating to the mechanism of 'voluntary' movements and it continues unchallenged in the latest editions of the text books. This doctrine holds that there are centres in the cerebral cortex where voluntary movements are represented. The evidence for this is that electrical stimulation of a small part of the surface of the cortex, known from now on as the motor gyrus, produces movements, and removal or disease of the excitable region produces a loss of voluntary movement. It is believed that from these cortical motor centres fibres carrying voluntary motor impulses run in the pyramidal or cortico-spinal tracts direct to the spinal cord to activate the necessary motor cells in the anterior horn.

Now this simple doctrine conceals serious problems of verbal definition, but the only aspect I shall discuss centres on the inferences we should draw from certain well-known facts of observation. The principal difficulty here lies in resisting the temptation to be over-enthusiastic about what the techniques can offer.

The first question to be faced is 'what should we infer from the observation that stimulation of a strip of cortex produces movement?' The most restrained interpretation would be that a response to stimulation merely indicates some continuity of nervous conduction between a region in the vicinity of the electrodes and a region in the vicinity of certain spinal motor cells. It does not imply necessarily that only one pathway is involved, nor does it imply how many synaptic junctions there may be. In particular we cannot assume to be stimulating that fraction of

the cortico-spinal pathway subserved by the Betz cell population of the precentral gyrus, although many workers make this inference. In other words, Ferrier, in 1873, was doing the first experiment in this country in electrical neuronography. He was giving evidence for a new conducting pathway and three years later quite independently Flechsig demonstrated a complex and variable collection of descending fibres now known as the cortico-spinal pathway. But there is still no evidence that all the phenomena depend directly on the pattern of impulses in this pathway. In fact there is some suggestion to the contrary, for the movements may depend on the position of the limbs and the point at which this afferent information acts is probably spinal.

The second question is ' Because stimulation of the cortex produces movement can you infer that you have demonstrated centres for movement in the grey matter ? ' This was Ferrier's inference, but a few months after his first paper appeared, Burdon Sanderson showed that all the movements that could be produced by stimulation of the grey matter could be produced by stimulation of the white fibres after the grey matter had been cut away. In observations in Man, Penfield has confirmed Burdon Sanderson's results and so have I. So far as the detailed character of the movement is concerned, stimulation of the white fibres and of the grey matter gives identical movements in the same individual. Without the grey matter the latent period between stimulus and response is usually shorter. It is not difficult to obtain the smallest discrete movements, nor if you look more closely as Sherrington did, inhibition of contraction from stimulation of the fibres as they pass through the basal ganglia.

In other words, from the nature of the movements produced it is sufficient to assume that a cortical stimulus is affecting only the final efferent neurone in whatever cortical chain there may be. There is no evidence from stimulation for a type of representation in the grey matter more complex than is represented at the final cortical efferent neurone.

It has for some years been implied that alteration of the character of the movement with altered frequency of stimulation was a special property of the grey matter, but in a recent paper Rasmussen and Chusid have stated that if such an alteration can be obtained from a surface point it can equally well be obtained from the white matter beneath if the grey matter is removed.

It seems that the only special properties of cortical grey matter revealed by stimulation are those associated with the latency of the response, with secondary facilitation and with clonic ' after discharge ' movements ; and although the grey matter doubtless has a lot to do under normal conditions we must be guarded in the inferences we make from stimulating it. The extent to which an electrical stimulus can be referred to as physiological will be discussed later but we should remember that the grey matter requires relatively strong stimuli to give responses. For example, in a conscious man, to produce a small movement or sensation in the fingers by cortical stimulation requires a strength of stimulus which if applied instead to the tip of his tongue would be definitely pungent. The need for care is further emphasised if we ask a third question : ' Do the movements in response to stimulation, by their particular character or quality suggest the inference that the area is concerned with a particular sort of movement ? ' Ferrier interpreted the movements as characteristically purposive but it was Sherrington who first took an opposite view. ' It is their fractional character which makes equivocal any purpose that an observer, who would interpret their purpose, can assign to them.' Penfield has frequently stressed the essentially primitive, athetoid or infantile character of cortically induced movements in the conscious human subject. My own observations agree with this, but if we face these facts of observation and yet somehow identify the cortex with the elaboration of complex movements, the only way to reconcile observation and belief is to stress that stimulation is unphysiological and to follow Sherrington in the inference that the cortex builds up elaborate movement patterns from the elementary bricks which this unphysiological assault reveals. I think there are two difficulties in this line of reasoning. In the first place, clinical neurologists

would be hard put to bring forward evidence from man to give strong support to Sherrington's scheme. It is, of course, the case that patients with cerebral disease occasionally, not often, present a specific loss of skilled and complex movements, but as Macdonald Critchley has recently stressed this may be a reflection of disease of many other parts of the brain, not necessarily of the cerebral hemisphere. In the second place, ideas, memories and emotional changes may result from electrical stimulation that are as real as when they arise without it. One should hesitate to suggest that a stimulus had a physiological refinement in one part of the cortex which it conveniently lacked in another. The point is I think that electrical stimulation of the cortex produces a disorganisation of a type indistinguishable from that which occurs in some forms of epilepsy. We would probably be better to drop the word stimulation altogether when we refer to the cortex and refer to the technique as electrical disorganisation. But the evidence from epilepsy suggests that this disorganisation is within physiological limits in the sense that the symptoms may result from temporary alteration in the behaviour of a normal gyrus. To refer to the technique as unphysiological does not permit one to discard the observations because the movements are too gross to be convenient for a particular hypothesis, since on a neighbouring gyrus the same stimulator may produce a response of vast complexity and refinement.

The significance of the responses to stimulation of the motor cortex lies rather in the fact that so far as the movements can be said to show organisation there is little if any evidence that in animals they show more than is shown by dorsal root stimulation. It was Sherrington in particular who stressed the similarity between cortical movements and spinal movements in the same animal, particularly remarking that opposition of the thumb in the monkey was also the movement most frequently and easily obtained by dorsal root stimulation. This is a very relevant observation because, following statements by Huxley and Duchenne to the effect that the thumb was the most voluntary part of the body, the ease with which movement of it was produced by cortical stimulation has been held to be good evidence for the role of the cortex in voluntary movement. I think in short that in our conservative estimate of what stimulation tells us, we should regard both the character of the movements and their distribution through the musculature as indicative of no more than the accessibility of part of the cortex to motor organisations at a spinal level.

If we consider now ablation, it is tempting to assume that stimulation and ablation are complementary techniques when applied to the cortex just as they are, for example, in a peripheral motor nerve. The paralysis that may follow ablation is thus presumed to be accounted for by the removal of motor centres which stimulation has revealed. But is not this too simple a view? The fact is that the results of cortical stimulation and ablation may in a sense be contradictory. I will give three examples : First we know from observations by Penfield and others in man, that stimulation of certain points of the brain in the conscious human subject may disturb in various ways the act of speaking. Yet excision of one of these points is not necessarily followed by any impairment of speech in the patient. In the second place it is now known that there is an extensive region on the medial surface of the hemisphere which when stimulated gives a motor map, and excision of this excitable region or its disruption by disease in man is, so far as I know, never followed by detectable loss of movement. In the third place some recent evidence from stimulation of the classical motor region in cases of long-standing disease in man has clearly shown that in these cases a classical motor map can sometimes be obtained, but when the whole hemisphere is removed there is not necessarily any loss of those movements produced by stimulation—it is not that function recovers, but it is never further disturbed by removal of excitable tissue.

If these contradictions between stimulation and ablation appear to conflict with our ideas then this state is of our own making and we should be encouraged to look anew at the foundations of the classical doctrine of voluntary movement because this rests on the complimentary

evidence of the two techniques. I suggest we follow a line of thought suggested first by Francois Franck in 1886. The idea is that we should cease to regard the cortico-spinal tract as an efferent tract from an area where movements are represented and regard it instead as an afferent tract to a region where they are represented. To help you to adopt this unfamiliar view I would ask you to imagine the nervous system from the point of view of the anterior horn cell of the spinal cord. From that position looking backwards into the grey matter you see among other things various fibres terminating in the space some way behind you. You know that an impulse in some of these fibres will affect the pattern in which you and your neighbours are depolarised, in the same way. One set comes, you are told, from the outer surface of the body through the dorsal root, another less defined set has components connected with the outer surface of the cerebral cortex via the medullary pyramid. To you, both these sets are equally afferent—your own axon is efferent. Therefore, you take a common-sense view of the situation and hold that it is no more significant to say that movements are represented in the cortex than it is to say that movements are represented in the skin.

If we adopt this somewhat altered view of the motor system I think we must make use of a particular hypothesis before it becomes of any use. The hypothesis is that it is a general property of the motor system that interference with the function of any two of its component elements gives more disturbance than complete elimination of one single component. There are certain observations which make this rather vague idea not improbable, but the difficulty in substantiating it lies in the fact that it is impossible with present techniques to observe the effects of destruction of one single component. But I would expect that destruction strictly limited to the group of crossed cortico-spinal fibres and nothing else at all, would give no great interference with function. This is not quite such a bold supposition as it may seem because surgeons, on and off for many years, have been cutting the lateral columns in the high cervical region of the cord, and thereby interrupting we must presume a considerable proportion of these fibres, and they have reported that there is surprisingly little impairment of fine movement or, indeed, any disability at all save for a slight overall weakness of movement in the hand.

There is other evidence that the conspicuous cortico-spinal tract has been given too much significance. It has not the unique over-development in primates that it was once supposed to have. It is comparatively well developed in hoofed animals, and in the rat. Lashley 30 years ago reminded us that this tract occupies a proportionately larger area of the cross-section of the medulla than it does in man. It has never been convincing to identify it with specially discrete or delicate movement *per se*. Looking down the scale of complexity from birds, crabs to spiders, one sees that delicate and graded movement can occur in particular instances without the most elaborate nervous equipment. The increase in bulk of the tract may be related to the increase in the number of separate and varied types of movement an animal possesses, though no one seems to have suggested a way of enumerating these quantitively.

Turning now to the accepted results of cortical stimulation and ablation, which in some situations are apparently complementary in their support for a belief in cortical motor centres, in others contradictory, I suggest first that we regard these techniques as primarily experiments in topographical anatomy rather than physiology. And I suggest we consider it to be the general rule that the two techniques when applied together to a given cortical area are *not* complementary. It may be that the results of stimulation depend in a sense on a misleading anatomical peculiarity which has provided one exposed surface from which one particular lot of fibres run in an afferent sense to the only motor centres of the body, those of the spinal cord, the one place where all possible information is represented—in this centre the distinction between efferent and afferent is largely meaningless, all fibres are equally afferent. It is premature to identify any one of the afferents with a particular type of movement in the intact animal, particularly if we use private introspective data as the basis, as we do

implicitly in the use of the term 'voluntary.' We should by now be free from the earlier naive mechanistic belief that the cortex is the seat of the will—a belief which naturally forced attention on a direct cortico-spinal pathway. It may be that stimulation of the cortex in general does not demonstrate that fibres are efferent from a functioning motor centre, but rather that in the region of the electrode there is the origin of one of the many sets of afferents arising from perhaps a dozen or more different regions on both sides of the hind, middle and fore brains to impinge on each spinal centre. If it is correct to regard functional loss as indicating to us the simultaneous loss of two or more afferent systems, then we shall hold the loss following motor cortex ablation in normal creatures to result from the involvement of two or more systems. The motor cortex is seen to be unique in one respect, that disturbance of it involves simultaneously two at least, probably more, of the systems having afferent connection with the spinal motor centres, and the apparent concurrence between the evidence from stimulation and ablation of this complex region is misleading. The contradictory effects of stimulation and ablation in normal creatures to which I referred—the cases of the speech and supplementary motor areas—may be exemplifying the general rule that from a single excitable cortical focus there arises only one afferent system whose loss can be covered up. In the cases of long-standing disease of the hemisphere where an excitable motor cortex can be removed without further deficit, we see a patient with involvement already of many afferent elements to the spinal motor centres having direct and indirect connection as they do with many parts of the cortex. The diseased limb is presumably influenced by intact lower mechanisms and surgical removal of the proportionately small cortical component passes unnoticed.

Sensori-motor Skill in Monkeys

by

Dr. F. P. Glees and J. Cole

In our film we showed three tests we have developed to measure sensory and motor ability in monkeys. The tests which are fully described by Cole (1952), have been successfully used by us in studying the effect of small lesions in the motor cortex (Brodmann's area 4) [Glees and Cole (1950)], and in our work on the sensory cortex (areas 1, 3, and 2) [Cole and Glees (1954)].

We have been able to measure the impairment resulting from small lesions in hand region of both the motor and sensory areas, and to record by daily testing the gradual recovery of motor power and sensory ability.

The first test is a simple match-box drawer with a protruding lip, which the monkey grips between thumb and fore-finger, and pulls the drawer open against a variable weight to obtain a reward from the drawer. This test enables us to estimate the strength of the grip between thumb and fore-finger.

The second test measures dexterity in the use of the same digits; it is a flat board 20 × 15 cm. with twelve circular hollows $\frac{1}{2}$ cm. deep and $2\frac{1}{2}$ cm. in diameter. In each hollow is placed a cube of apple or carrot and the time taken to pick up all of the cubes is recorded by a stop-watch.

The third test is for tactile discrimination of three-dimensional solids, and consists of a cone and a pyramid $5\frac{1}{2}$ cm. in height, which the monkey is taught to discriminate between by touch when they are placed in a deep leather bag. Choice of the positive object always being rewarded.

Lesions in the sensory cortex produce impairment in the performance of all three tests, i.e. loss of motor power, dexterity and stereognosis, but when the lesion is small, recovery always takes place, and in some cases a pre-operative level of performance is achieved.

Macaca mulatta and *M. nemestrina* monkeys readily learn the first two tests, but so far we have only been able to train *M. nemestrina* monkeys to the third test.

The tests shown in our film and described in this paper, lead us to consider how far these tests and our experimental work have advanced our knowledge of cortical localisation.

Diagrams in textbooks of anatomy and physiology illustrating the connections

of the motor and sensory cortex, have frequently misrepresented by gross oversimplification, if not falsification, what is in truth a highly complex arrangement of nervous pathways about which little is really known.

The downward connections of the motor cortex to the spinal cord are said to run in a neat arrangement of parallel fibres coming from the appropriate cortical subdivision. However, if the embryological conditions under which this outgrowth of cortical axons occurs is considered, it is highly unlikely that some kind of 'traffic police' could exist during development, which would direct fibres entering the spinal cord to their appropriate segmental level.

We tested the classical assumption that strictly defined pathways pass from the motor cortex to the spinal cord, by causing small discrete lesions in the motor cortex and then studying the resulting Marchi degeneration. From these studies, there was no doubt that fibres intermingle in their descending course and change their positions all the time. There was neither evidence of a localised area of fibres of a cortical subdivision at medullary level, nor of any lamination of cortico-spinal fibres at cord level.

We would therefore suggest that cortical connections at spinal cord level are established largely by functional selection, and are not formed by a mysterious 'anatomical soldering' process which brings cortical axons in contact with spinal neurons. This functional selection which we consider must occur in late foetal life and in infancy, from fibres arranged at random, would eventually lead to the temporary disuse or even destruction of certain collaterals of this random arrangement. When this process of maturation has been completed, ablation of a subdivision of the motor cortex results in muscular weakness, but this is only of a temporary nature and considerable recovery occurs due to reorganisation and the use of pathways still unimpaired. Certain observations are more easily explained by this flexible conception of motor pathways. For instance movements normally associated with specific parts of the body, e.g. the fingers, can in cases of congenital abnormality or amputation in man, be replaced by another part of the body, e.g. the toes. Or loss of sight can be replaced in reading by use of the finger tips as in Braille, and as Macdonald Critchley (1953) has shown, even the tip of the nose can equally well be used by blind people.

We considered also the possibility that the organisation of the motor system might be governed (influenced) by an orderly system existing in the sensory cortex and its afferent channels. However, although studies of sensory tracts projecting on to the sensory cortex show an orderly arrangement up to the posterior column nuclei, it is highly improbable that this rigid order continues within the medial fillet (which is the continuation of this system). Furthermore, at thalamic level a considerable overlap exists, which must become more complex during the embryological stage, when thalamic neurons grow through the internal capsule in order to terminate within the sensory cortex.

Ablation of the sensory subdivision of the post-central gyrus results in marked functional loss, but recovery occurs, greatly helped by the training shown in our film, which assists the animal to overcome its initial impairment. This recovery, which has been tested, can best be explained by the overlapping of the sensory pathways.

We consider from our studies that only *gross* cortical localisation results from anatomical arrangement of the motor and sensory fibre tracts. Whereas *fine* cortical localisation seems to be determined by the influence of innate or environmental factors upon the relatively unspecific anatomical arrangement of fibre connections. Loss of cerebral cortex can thus be compensated for by the remaining cortex taking over fibre connections hitherto unselected, and thereby establishing useful function.

References

Glees, P. and Cole, J. (1950) : 'Recovery of Skilled Motor Function after small repeated Lesions of Motor Cortex in Macaque.' *J. Neurophysiol.* **13,** 137–148.

Cole, J. (1952) : 'Three Tests for the Study of Motor and Sensory abilities in the Monkeys.' *J. comp. physiol. Psychol.* **45,** 226–230.

Critchley, M. (1953) : "Tactile Thought, with special reference to the Blind.' *Brain,* **76,** 19–35.

Cole, J. and Glees, P. (1954) : 'Effects of small Lesions in the Sensory Cortex in Trained Monkeys.' *J. Neurophysiol.* **17,** 1–13.

Lesions of the Cerebral Hemispheres and their Effects upon Optokinetic and Caloric Nystagmus

by

Dr. E. A. Carmichael, Dr. M. R. Dix and Dr. C. S. Hallpike

Introduction

In 1923 De Barenne and De Kleyn showed that removal of one cerebral hemisphere in the rabbit made it easier to induce vestibular nystagmus with its rapid component towards the affected side than in the opposite direction. In other words, there was a directional preponderance of induced vestibular nystagmus to the side of the lesion. In 1942 Fitzgerald and Hallpike were able to confirm and extend these conclusions in the course of some observations made upon human subjects suffering from a variety of unilateral cerebral lesions. The caloric method of excitation was used with standardisation of the stimuli, and the results were recorded in terms of the duration of the resultant nystagmus.

The substance of their findings with human cerebral lesions was that the caloric responses were normal if the lesion did not affect the temporal lobe. If, however, it did, then caloric nystagmus showed a preponderance towards the affected side.

With regard to the neurological mechanism of this phenomenon an explanation was suggested based upon the experimental work of De Kleyn, who had shown that the primary reflex arc of vestibular nystagmus comprised the vestibules, vestibular nerves, vestibular nuclei and a short pathway within the brain stem to the eye muscle nuclei. If this were preserved, then vestibular nystagmus could be readily induced in the absence of the cerebral hemispheres. The effect of the hemispheres, or rather the temporal lobes upon this nystagmus was, therefore, rightly to be described as a modifying influence.

Since, as they pointed out, the chief cell stations in the reflex circuits for nystagmus to the right and left were respectively the right and left vestibular nuclei, Fitzgerald and Hallpike accordingly suggested that the effect of the temporal lobes in modifying vestibular nystagmus was exerted by means of cortico-fugal fibres passing to these nuclei, each temporal lobe being associated in this way with one set of nuclei. Thus, destruction of one temporal lobe would result in a corresponding directional preponderance of caloric nystagmus.

Present Investigation

This has been carried out in a larger series of subjects with a wide variety of supratentorial lesions, with a view to confirming and extending this finding, that temporal lobe lesions bring about a directional preponderance of caloric nystagmus to the affected side.

In addition, we have endeavoured to throw light upon the closely associated problem of the cerebral control of optokinetic nystagmus. What is known of its neurology can be stated briefly as follows :

Animal experimental studies, in particular those of Smith and Bridgman, make it evident that optokinetic nystagmus is primarily a brain stem reflex dependent upon direct nervous pathways from the optic tracts to the superior colliculi and eye muscle nuclei. If these are preserved, then optokinetic nystagmus may be induced following removal of the hemispheres, provided only that the visual stimulus involves a substantial area of the retina. If, however, the stimulus is a small one presented upon a large, stationary background, then in most animals a nystagmic response can be obtained only if the cerebral cortex is present. With the primary brain stem reflexes we have not been concerned in the course of the present investigations. We have studied instead the part played by the cerebral hemispheres. In particular we have endeavoured to identify the parts of the hemispheres which are concerned with the control of optokinetic nystagmus, and the way in which they exert this control.

Material

This has consisted of 73 subjects, all in-patients at the National Hospital, Queen Square. All had undergone full neurological and otological examination, and in every one a unilateral cerebral lesion had been localised. The great majority of these subjects have been divided, for purpose of analysis, into four groups based upon the results of the

caloric and optokinetic tests. This analysis is shown in the following Table :

DIRECTIONAL PREPONDERANCE

	Caloric Nystagmus	*Optokinetic Nystagmus*
GROUP 1		
34 subjects	Present	Present
GROUP 2		
19 subjects	Absent	Absent
GROUP 3		
13 subjects	Present	Absent
GROUP 4		
3 subjects	Absent or Slight	Present

Group I. Twenty-seven of our first group of thirty-four patients suffered from infantile hemiplegia, and in fifteen of these it was possible to establish the site and nature of the pathological process by examination of the affected hemisphere removed at operation. Two types of pathology were found. In nine of the fifteen, large cysts were present with disorganisation of the brain over the territory of the middle cerebral artery. In the other six, diffuse areas of destruction, based in the main upon occlusion of small vessels, were present throughout the hemisphere. If the pathological findings in these fifteen ablated hemispheres be applied to all of our twenty-seven hemiplegic patients, then it seems clear that of all of them we can say that they suffered from a widespread structural disorder of one hemisphere, with some concentration of the destruction in the cystic cases to the territory of the middle cerebral artery.

Our findings in these twenty-seven cases make it clear that these two physical signs, if so we may call them, directional preponderance of caloric and optokinetic nystagmus, are constant accompaniments of any severe, widespread damage of one hemisphere.

As to the localisation of particular areas of damage to which these signs can be attributed, little can be said on account of the diffuse nature of the damage. Much more helpful were our findings in the remaining seven cases of this group, in which more circumscribed lesions were present : meningiomata, gliomata and others.

An analysis of the position and extent of these lesions showed that the zone of destruction common to all was the posterior temporo-parietal zone of the cortex.

This information, therefore, supplements that provided by our hemiplegic group, and shows that within this zone lie the controlling centres for both caloric and optokinetic nystagmus.

Group II. In our next group of nineteen cases, directional preponderance of caloric and optokinetic nystagmus was absent. The lesions were scattered about the hemispheres, being confined to certain clearly defined cortical areas. These were the frontal lobes, the anterior parts of the temporal lobes, and the upper parts of the parietal lobes. None of the nineteen lesions of this group caused directional preponderance, and the findings in this group are, therefore, complementary to those obtained in Group I, in which both varieties of directional preponderance were present. Thus, in Group I, lesions of the posterior temporo-parietal area of the cortex were present in all cases. In Group II this area was unaffected.

Group III. The findings in our third group of thirteen subjects who exhibited only a directional preponderance of caloric nystagmus have made possible an important step in our analysis of this posterior temporo-parietal zone.

Thus, in all thirteen cases the lesions were confined to the posterior part of the temporal lobe, and this, therefore, is shown to be the site of the cerebral controlling centre for caloric nystagmus.

Group IV. In this group directional preponderance of optokinetic nystagmus was present with very slight or absent directional preponderance of caloric nystagmus. The material in this group has been much scantier and more difficult to analyse. In three subjects lesions were present in or near to the calcarine fissure. In none was there any disturbance of caloric nystagmus. In one case optokinetic nystagmus was normal. In the other two cases directional preponderance of optokinetic nystagmus was present although of very slight degree.

We have, therefore, concluded that lesions of this part of the cortex play no part in the directional control of caloric nystagmus. Optokinetic nystagmus may be affected, but only to a small extent.

In four other subjects lesions were

present on the outer surface of the hemisphere involving the cortex around the postero-inferior part of the parietal lobe. One of the subjects was examined three weeks before death. The outstanding findings were a homonymous hemianopia, a slight directional preponderance of caloric nystagmus, and a very severe directional preponderance of optokinetic nystagmus. At post mortem there was gross distortion by tumour tissue of the supramarginal and angular gyri.

The other three subjects all had lesions in the neighbourhood of, but not directly involving, the supramarginal and angular gyri, and in the course of operative explorations incisions were carried into these gyri. Directional preponderance of optokinetic nystagmus was absent or very slight before these operations, but became very marked following them.

It would, therefore, seem justifiable to conclude that interference with these gyri in the course of these operations brought about a selective increase of directional preponderance of optokinetic nystagmus.

On the whole, therefore, we think that the findings in our fourth group of cases provide a strong indication that the cerebral centres which are concerned with the directional control of optokinetic nystagmus are situated in the supramarginal and angular gyri.

To sum up our findings, large lesions of the posterior temporo-parietal region cause directional preponderance to the affected side, both of caloric and optokinetic nystagmus.

The clinical finding that either of these derangements may occur independently, suggests that separate centres exist in this area for the directional control of these two varieties of nystagmus.

This suggestion is confirmed by our pathological findings in cases in which one or other of these derangements occurs alone. Thus, the area concerned with caloric nystagmus is shown to lie in the posterior part of the temporal lobe, while that concerned with optokinetic nystagmus lies in the supramarginal and angular gyri.

THE ASSOCIATION OF PSYCHOLOGY WITH OTHER DISCIPLINES[1]

by

PROF. F. V. SMITH

THE discussion of some possible affiliations with Psychology directs attention to the scope of Psychology and this is probably more reliably indicated by what psychologists are doing or have done, by the content of papers appearing in psychological and related journals, than by the more formal considerations which would prescribe what Psychology ought to be. The well-established journal *Psychological Abstracts* publishes abstracts from a list of over 400 journals, the criteria of selection being that the articles are thought to be relevant to the psychologists' activities. The journals include many which would be regarded as directly serving the disciplines of Zoology, Neurology, Physiology, Neurophysiology, Pharmacology, Anthropology, Encephalography, Sociology, Education and Philosophy and technical advances in the recording of human reactions which depend upon the several branches of Physics, also appear. The exacting problems encountered in Psychometrics or psychological measurement also require reference to and have more than once provoked advances in what might be termed ' pure ' statistics.

It would thus appear perfectly defensible to define the scope of the psychologists' *interests* as the study of animal behaviour, where the term ' study ' includes both description and attempts at explanation of the behaviour, and the term animal covers the range from the unicellular organisms to the behaviour of human beings, both normal and abnormal. Obviously, anyone interested in the processes of human learning and adjustment can with advantage examine the work done on the learning and adjustment of organisms having simpler nervous systems. So, too, the studies of adjustment to the demands of the group in the so-called ' primitive ' societies are of value in the study of similar problems in the so-called ' civilised ' societies. Effective and defensible incentive schemes in different phases of industry involve considerations of interest for anyone concerned to explain behaviour as do the aspirations and frustrations arising within the so-called ' primitive ' societies. Anyone interested in the psychology of thinking, though he may not find the solution to many of the problems in this field, tends in these times to be attracted to the literature of Encephalography and the frequently confused but provocative literature of Cybernetics.

Possibly more than any other discipline to-day, Psychology demands reference to other subjects even though the central problems in Psychology, such as the determination of the criteria of the satisfactory explanation of psychological processes are of especial difficulty and peculiar to the subject. Yet clearly the psychologist cannot be expected to be omniscient. The scope of his interests must cover anything relevant to that aspect of behaviour which is his special interest ; but this is to define the *total* scope including the penumbra around the central core of interest with which all those engaged in the subject are ultimately concerned namely, human behaviour. Because contributions to behaviour covered by the literature of many of the subjects mentioned above are conceivable, it is understandable that reference should be made to them and at the present time there are perhaps additional reasons for this. Many of the questions addressed to Psychology to-day are in the form of those addressed to a natural science. Parents and other responsible

[1] Abstract of a paper delivered to Section J (Psychology) on September 7, 1954, at the Oxford Meeting of the British Association.

persons, to take but one example, while no doubt interested in the incidence and description of delinquent behaviour, want to know *why* a particular child is delinquent or backward and the implication is always that if an aetiology could be found, some remedial or preventive treatment might be possible. Such an aetiology could involve a reference to the vast literature on motivation, a field in which there is much further work to be done, to the previous history of the child, to the geographer's study of the movements of industrial population and the decline in the use and social status of a particular area (8), to the educational psychologist's analysis of the child's abilities relative to a particular educational system, to a full report of the psychological tensions and the physical conditions within the home, as well as to the child's general health.

Historically, the issue of whether Psychology is a descriptive rather than a natural or explanatory science appears to have been most discussed in association with the writings of Wilhelm Dilthey (4). Briefly, Dilthey's view was that an explanatory science endeavours to go beyond the data immediately given in order to account for the recorded observations. To this end, hypotheses and constructs, very often associated with unobservable features, are employed and an ordered account of events 'out there' is achieved. The natural sciences are also associated with the spatial and temporal qualities of perception of the external world and their discrete character; but for Dilthey, experience of psychological states in ourselves and to a large extent in others, has no such clearly defined units. The associationism of Hume, Hartley and Mill, for example, he regarded as a mistaken attempt to follow the model of the physical sciences. Attempts to go beyond the unity and coherence of inner experience he regarded as enforcing the procedures of the physical sciences, where coherence is not immediately apparent, upon psychological processes, where the coherence and organisation of mind is given and is evident in perception. He pointed out that literary studies of personalities, based upon direct experience of the persons concerned over appreciable periods of time, very often achieved a degree of understanding and insight which was greater than any of the 'psycho-physical' methods known to him at the close of the nineteenth century. The method of direct experience of personalities was thus preferred and on this basis regularities or consistencies could be recorded. In consequence Psychology was regarded as a descriptive science rather than an explanatory science.

Every science has its descriptive phases, when observations are recorded and classified. The issue is whether or not Psychology should remain exclusively in the descriptive phase and attempt nothing further than the discernment of consistent features, and here the following considerations appear relevant:

(i) In Dilthey's view explanatory psychology was unable to do more than conjoin nodal points in the complex and changing processes of mental life. Bergson (2), too, drew attention to the cumulative nature of experience, such that each successive experience is interpreted from a cumulative and thus different background. In this way, one cannot have precisely the same experience twice and in view of the inevitable differences in the experience of different persons, the probability of a given external situation representing precisely the same experience for any two persons is small. It is nevertheless true that if the requirements of precision are raised sufficiently high, the probability, even in the physical sciences, of precisely the same event recurring, would be small. There is, too, in the investigation of many events in the physical world a stage when conjunctions of nodal or outstanding features is all that can be accomplished. At a later stage finer discriminations can be made and functions relating the several variables involved in the phenomenon can be indicated. From one important aspect of methodology therefore, the difference between physical and psychological processes is one of degree.

(ii) Writers since Dilthey, including G. W. Allport (1), have emphasised the individual or idiographic nature of each personality. Dilthey believed that the several aspects of the personality could be best comprehended in a 'verstehende' situation. Something akin to this is probably achieved by what is now called

' clinical insight.' The clinician concludes that he has comprehended the salient features of the personality, their relative importance and in association with these features, the significance of particular incidents in the patient's life. Such conclusions, very often amounting to conviction upon which action is based, do presume a general fund or background of experience of personalities, a measure of consistency or generality in the phenomena associated with personality. The issue is whether or not the degree of consistency in the experience of these phenomena is sufficient to justify the later stages which are more typical in the explanatory sciences, namely the formulation of hypotheses, deduction of consequences and the rigorous testing of hypotheses and deductions against factual observation. It is not a valid objection to claim that human behaviour and personalities are too complex and that the practical difficulties of discerning conditions which are spatially and temporally continuous with the behaviour of the person and functionally related to it are considerable. The crucial issue is whether or not such procedures are in principle applicable and have yielded useful results for some aspects of human behaviour.

(iii) As Ebbinghaus (5) maintained, Dilthey's position assumed that the methods then current in Psychology could not be improved. In point of fact development in the first half of the twentieth century included a number of hypotheses which attempted to account for observations recorded. Naturally there is a profit and loss account including many hypotheses discarded or greatly modified ; but it may be fairly claimed that in more than one field the aetiology or explanation of many observed events is much clearer than at the time Dilthey wrote. The position is analogous to that of a map in which some of the major features have been sketched and items of detail are beginning to appear in several places. Juvenile delinquency, accident-proneness, occupational maladjustment and educational maladjustment, and even authoritarian attitudes have all been shown to be conditions with which a number of demonstrable factors may be associated. It is not always possible to indicate with precision those conditions which are spatially and temporally continuous with the observed event and functionally related to it ; but with the advent of more rigorous statistical methods, the analysis and clarification of many forms of association is possible. Expressed very simply, our understanding of many aspects of behaviour has been enhanced by seeking a knowledge of the contributing causes or by the approach to behaviour as a problem to be treated in the manner of the natural or explanatory sciences.

Such an approach does influence the choice of subjects which might be considered relevant to the study of Psychology. Since everything that human beings do has some psychological interest, it is not possible in the space available to deal with all the potential affiliations ; but within the scope of university studies the following are conceivable.

Philosophy-Psychology

This combination was frequently studied in the period before the advances in experimental design had greatly extended the scope of Psychology ; but the association can still be justified. Reference to the general history of ideas, the history and the philosophy of science and developments in Logic is of great value to those who are working on the methodological problems peculiar to Psychology. Even if deficient in experimental method, the earlier contributions of philosophers, particularly in the fields of perception and the thought processes, are still relevant to current investigations and definite advantages can accrue from a combination of philosophical analysis and modern experimental design. Again in the application of Psychology within society very definite ethical issues arise and an awareness of the literature of Ethics and of Political Philosophy is at least no disadvantage to the psychologist.

On the other hand the psychological study of such activities as concept formation, explanation, perceiving, learning and remembering can provide philosophers with much useful data.

Mathematics-Psychology

The application of statistical methods to Psychology frequently involves prob-

lems of interpretation. To some extent these can be overcome by care in experimental design ; but there are occasions when an indication of the interpretation which a mathematician would regard as defensible would be very welcome. A person who combines a thorough grasp of psychological issues with mathematical insight and competence is rare, perhaps too rare for his potential services to be appreciated. As to the contribution from Psychology, one could mention the study of the development from early infancy of the concepts which are fundamental in Logic and Mathematics. The work of Professor Piaget (6) in this field is sufficiently stimulating to suggest that there are implications in the development of which a person with a background in Mathematics (including Logic) and Psychology would have definite advantages.

Physics-Psychology

A study of the several branches of Physics can render so much experience meaningful that it is difficult to overestimate the importance of this basic science for the teaching of many other sciences. Psychology is no exception, as anyone who has conducted laboratory courses with students who are ignorant of physical principles, can confirm. With the current development of Psychology, a knowledge of electricity, electronics, sound and light is virtually necessary to read much of the experimental literature and to design apparatus for research. There is something to be said for acquiring this knowledge in a co-ordinated rather than an *ad hoc* manner. The considerable literature in psychological publications on visual and auditory perception would on the other hand serve to deter anyone who tended to overweight the contribution of physical factors. Quite apart from the advantages of mutual facilitation and the enhanced insight resulting from the affiliation of these two subjects, it is apparent that an effective co-ordination of the two would cover an appreciable and important sector of human experience.

Zoology-Psychology

The danger of anthropomorphism is one against which zoologists and psychologists have been very adequately warned. At the same time, psychologists must be interested in any adjustments made to the environment which appear to have been purposive or in some way associated with the operation of mind ; and in view of the difficulty of studying human minds in operation, there is at least a prima facie case for studying behavioural adjustments by organisms with simpler nervous systems. There may even be disadvantages in confining attention unduly to the relatively complex system of the rat. Clearly, if lessons from the behaviour of animals are to be learned by psychologists, a grounding in Zoology and Genetics is highly desirable, if not essential. The implication of evolutionary continuity in McDougall's hypothetical account of the basic human motives cannot be dismissed lightly ; but much careful experimental work at several phylogenetic levels must be done if comparative studies are to contribute to our understanding of such innate responses to stimuli as may exist.

The advances in Ethology may be regarded as a response to this need. On the other hand, there is a considerable literature in psychological publications on perception and animal learning which is worthy of reference by all those interested in animal behaviour.

Geography-Psychology

This combination of subjects is not usually proposed and may therefore be treated at greater length. Recent psychological literature has emphasised the contribution which the ' culture pattern ' can make to the personality of the individual and to the behaviour of groups. It is, of course, not a one-sided relationship. Personalities can influence cultures. Many influences are, however, represented in the culture pattern. Historical, economic and geographical factors are readily discernible and in most cases do not operate in isolation ; but if attention is for the moment arbitrarily confined to Geography, the influence of climate and factors of physical geography upon attitudes and behaviour constitute an interesting study. Although reported at a literary level, Pierre Buliard's (3) accounts of many of the attitudes of Eskimos indicate the appreciable contribution from an exacting environment. A recent broadcast talk on the

Third Programme by Professor Billington, (North-Western University), entitled ' The Mountain Men ' revealed similar evidence of an exacting environment inducing attitudes of opportunism and even mysticism. On the other hand, the power of man to modify his environment is considerable. As the eminent geographer Vidal de la Blache (7) remarked, 'A geographical individuality does not result simply from geological and climatic conditions. . . . It is man who reveals a country's individuality by moulding it to his own use,' and the study of the attitudes, motives and abilities associated with such a moulding of the environment is one in which the literature and techniques of Psychology could make an appreciable contribution.

There are, in fact, several ways in which the activities of geographers and psychologists could be complementary. In any regional study the geographers, among other things, would presumably furnish data on the physical features, climate, natural resources, distribution and movements of population, activities or occupations of the people and communications within the region. From the resources of Psychology, data could be provided from the research literature and by investigation on the influence of humidity and temperature on different types of human performance, on the distribution of abilities and the assessment of educational products, on job analyses giving the approximate kind and degree of skills required in different occupations, on incentives and satisfaction in work, on industrial relationships and the analysis of ' within group ' and ' between group ' tensions, on the assessment of attitudes, prejudices and stereotypes, on the susceptibility of groups to rumour and factors influencing the passage of rumour. The combination of the resources of the two subjects could quite conceivably endow the regional study with greater human interest and indeed make it a much more penetrating and useful document.

There are still further affiliations which could be discussed, such as those with Anthropology, Physiology, Neurophysiology and Pharmacology, the chief argument being that the combination results in enhanced insight and scope of application for each member of the combination.

The practical issue is how best to ensure effective combinations. Interesting and useful men are not infrequently found who have come into their present subject from another field of study by pursuing a particular line of interest ; but perhaps the best results from combination could be achieved by conducting the student's education in such a way that he becomes accustomed to assess what each subject has to offer and to contemplate those problems where the resources of both subjects are helpful. Joint honours schools in Faculties of Science to the level of the first degree, which could embrace nearly all the subjects mentioned, would be an obvious method and one which would tend to overcome the disadvantages of premature specialisation.

References

(1) Allport, G. W. *Personality : An Individual Interpretation.* Constable, London, 1949.

(2) Bergson, H., particularly in *Time and Free Will* (trans.). London, 1910.

(3) Buliard, P. *Inuk.* Macmillan, 1952.

(4) Dilthey, W. ' Ideen über eine beschreibende und zergliedernde Psychologie ' (1894) and other papers—*Gesammelte Schriften,* V. Band., Teubner, Leipzig u. Berlin, 1924.

(5) Ebbinghaus, H. ' Über erklärende u. beschreibende Psychologie.' *Z. Psychol.,* 1895, **9,** 161–205.

(6) Paiget, J. *Traité de logique.* Colin, 1949.
Introduction a l'Epistémologie Génétique. Press Universitaires de France, 1949–50.
And a number of other works such as :
La genèse du nombre chez l'enfant (with A. Szeminska). Delachaux et Niestlé, 1941.
La représentation de l'espace chez l'enfant (with B. Inhelder). Presses Universitaires de France, 1948.
La genèse de l'idée de hazard chez l'enfant (with B. Inhelder). Presses Universitaires de France, 1951.

(7) Vidal de la Blache, P. *The Personality of France.* Christophers, London, 1928, p. 14. Quoted by Watson *op. cit.,* p. 466.

(8) Watson, J. W., in Taylor, G. (ed.). *Geography in the Twentieth Century.* Methuen, London, 1951.
Professor Watson's chapter is entitled ' The Sociological aspects of Geography.' Part II, ' The Social Geography of a City,' is a study of Hamilton, Ontario.

THE SALIENT GEOGRAPHICAL FEATURES OF ELECTRICITY PRODUCTION IN GREAT BRITAIN[1]

by

E. M. RAWSTRON

ALTHOUGH the first electricity generating station established in Britain for public supply was driven by water power,[2] well over 95 per cent. of the current produced since then (1881) has been generated at coal-fired stations. Notwithstanding considerable progress in developing hydro-electric power in Scotland and to a smaller extent in Wales, coal must continue to provide about 95 per cent. of the annual output until such a time as it may be superseded by atomic energy. Hence electricity generated from coal alone is the subject of this article.

THE PATTERN TO 1948

During the early development of electricity in Britain both production and consumption were restricted to densely peopled parts of the country. It was technically impossible to distribute current more than very short distances, and the rural areas and small towns were insufficient as markets to support the higher capital charges of both small-scale generating equipment and the disproportionately lengthy distribution lines that are characteristic in areas of dispersed demand. By the end of the First World War the technical obstacles had been largely removed. No longer was the density of population the only possible determinant of the distribution and location of power stations. The industry was ready to supply the less densely peopled areas.

To foster and co-ordinate the growth of the system of supply a supervisory body—the Electricity Commission—was appointed by the Government in 1919. One of the Commission's functions was to collect and publish detailed statistics concerning the industry. These were intended both to facilitate supervision and to promote rivalry which would lead to greater efficiency between the local monopolistic undertakings. Quite incidentally they make possible, too, a close geographical study—almost unique among British industries—of the development of electricity production from 1920 onwards. The technical liberation of the industry from its urban ties and the wider field thus opened to supply, prepare the stage for change in the pattern of electricity production. The detailed statistics provide the means for undertaking the study.

In the event little change affecting the geographical pattern took place for twenty-five years, even in plans for the future. Figs. 1, 2 and 3 illustrate the development of the pattern from 1927 to 1947.[3] Two obvious conclusions can be drawn from even a cursory glance at this group of maps. First the total amount produced shows a steady increase and secondly the distribution pattern remains similar throughout. Changes in local detail did take place such as the closing of old stations in town centres and the erection of new on the urban periphery, the extreme case

[1] Address delivered to Section E (Geography) on September 6, 1954, during the Oxford Meeting of the British Association.

[2] At Godalming, Surrey, in 1881, and using power derived from the river Wey to light the streets of the town. As there was no legal power at the time to break up the streets the cables were laid in the gutters.

[3] Each circle represents one generating station except in London where stations are grouped into one large circle. This applies also to Figs. 4 and 5. In North Wales, the western part of the Southern Uplands in Scotland and in the Scottish Highlands are the major hydro-electric stations which are shown on these maps for comparative purposes.

being in the Birmingham district, but these affected the general pattern very little. Nor did the opening up of rural areas radically affect the pattern. Indeed, towns and the relegation of others in the same category to peak-load working. The smaller towns and rural areas were supplied largely from the main 'Grid' and

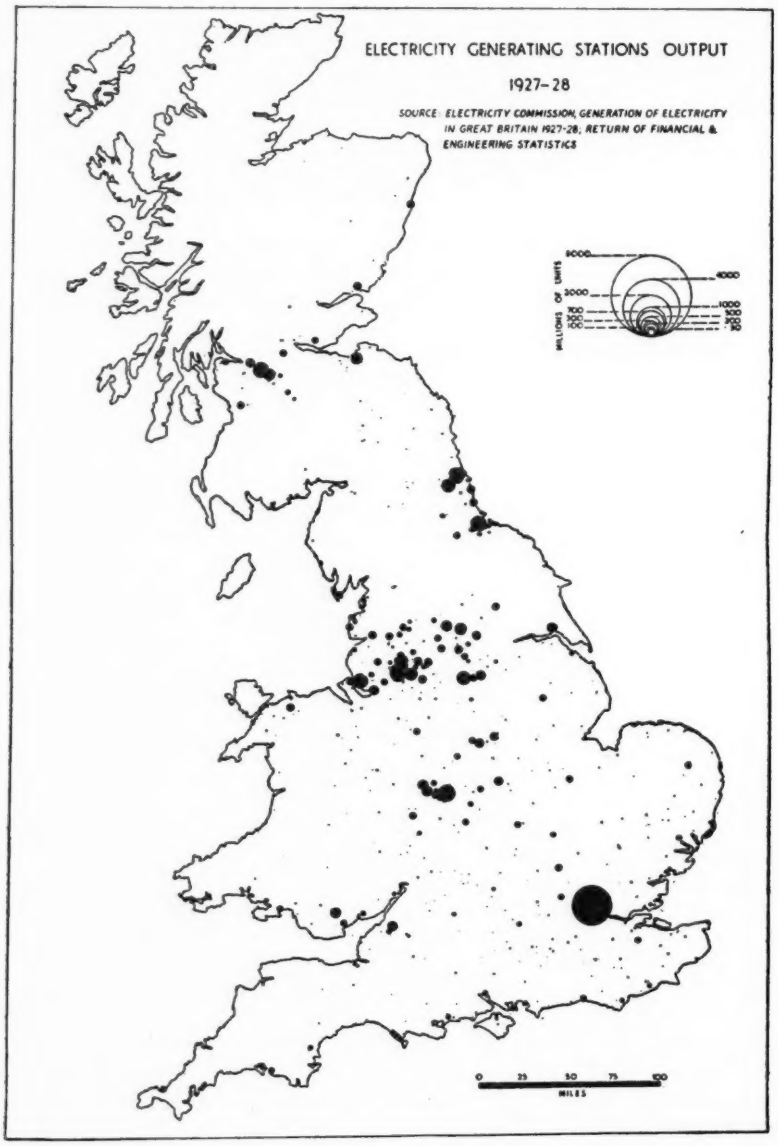


Fig. 1.

greater concentration not dispersion was rather the case.

The completion of an interconnection network—the 'Grid System'—between 1930 and 1936 resulted in the closure of many small stations supplying the smaller from its associated lower voltage distribution networks. In short the number of power stations declined during the period and by 1948 the remaining larger ones were almost all locationally related to some urban centre or group.

Nationalisation

On April 1st, 1948, the public electricity supply industry of Great Britain was nationalised and came under the control of two authorities. The North of Scotland Hydro-Electric Board became responsible for all generation, transmission and distribution roughly north of a line joining the estuary of the Tay and Loch Long. The area south of this line came under the control of the British Electricity Authority. It is this latter area with which the present article is concerned (Fig. 4), for there are only two coal-fired stations of importance in northern Scotland, those at Dundee and Aberdeen.

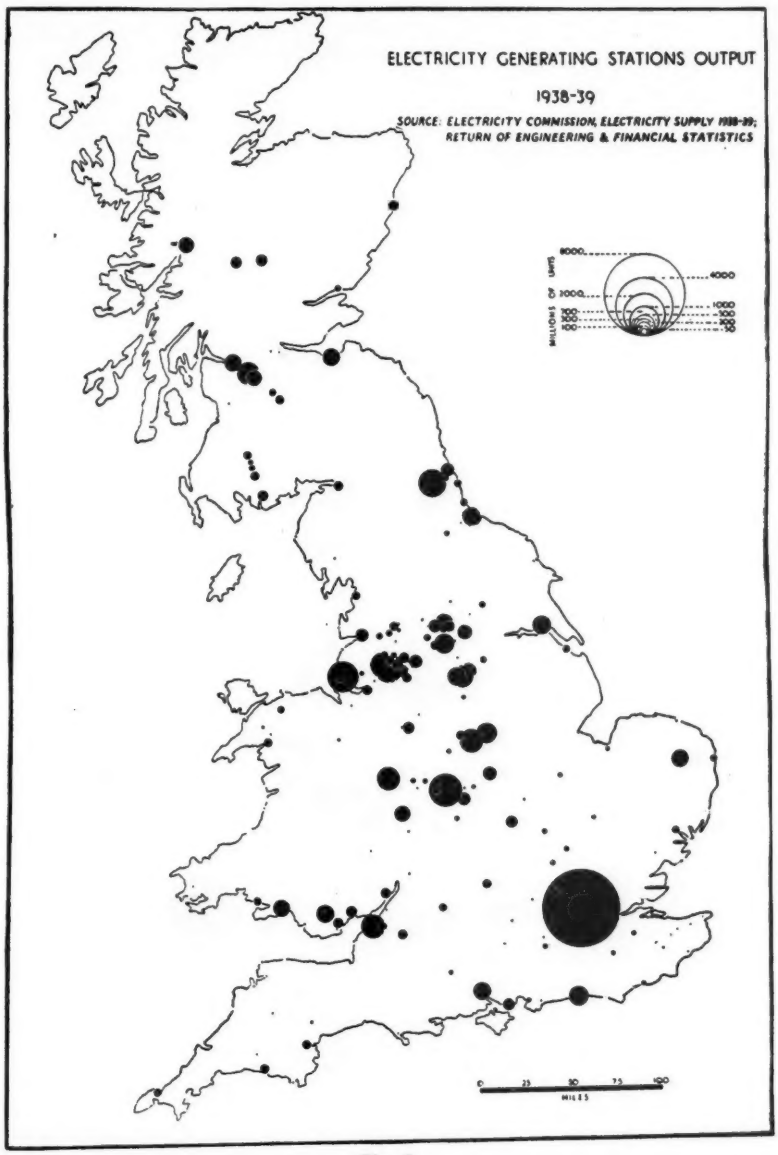


Fig. 2.

The Electricity Act (1947) which brought these changes marks a turning point not only in the organisation of the industry but in its economic geography too. The

geographical changes which are likely to ensue have certainly been greatly facilitated by nationalisation, but they would probably have taken place in any case under the pre-existing system of national supervision and control by the Electricity Commission and the Central Electricity Board (established 1926) provided certain features of the 1926 (Electricity) Act had been amended.

The Future Pattern

Statistics of output are not available for the future, so it is necessary at this point to change from mapping the output of stations to mapping their capacity. But Fig. 4, which shows capacity in 1948, is not vastly different in pattern from the maps showing output. The same concentrations appear and the prime relationship to population density is just as clear. The

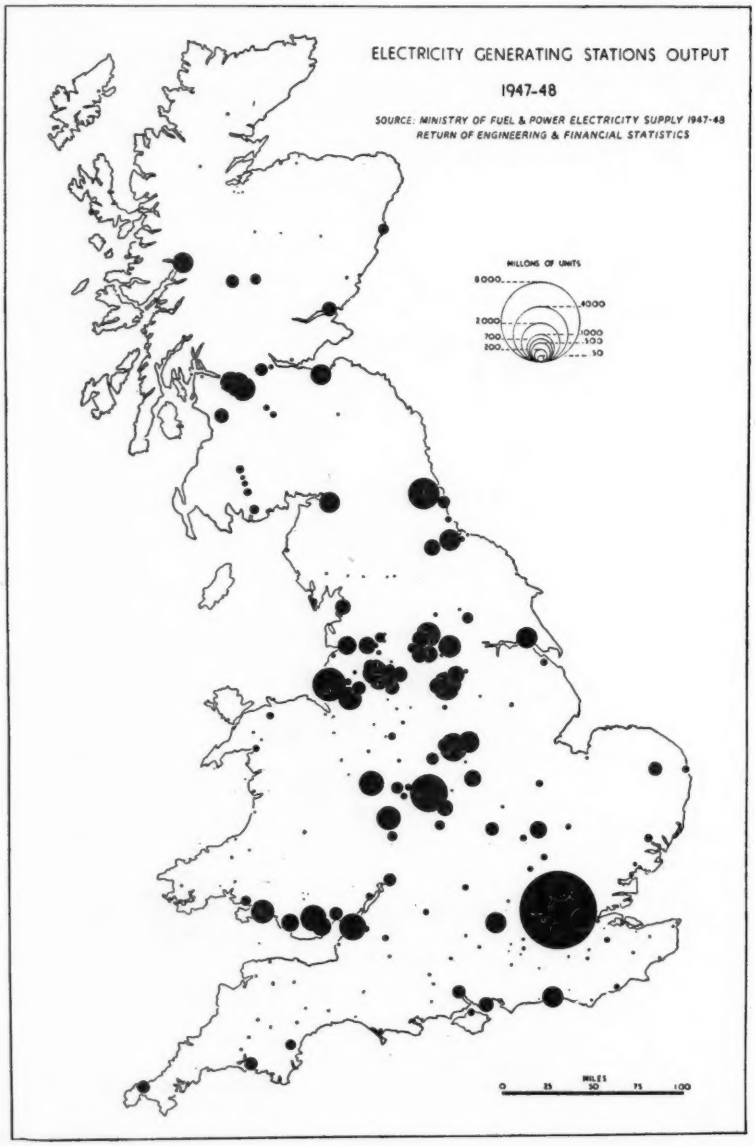


Fig. 3.

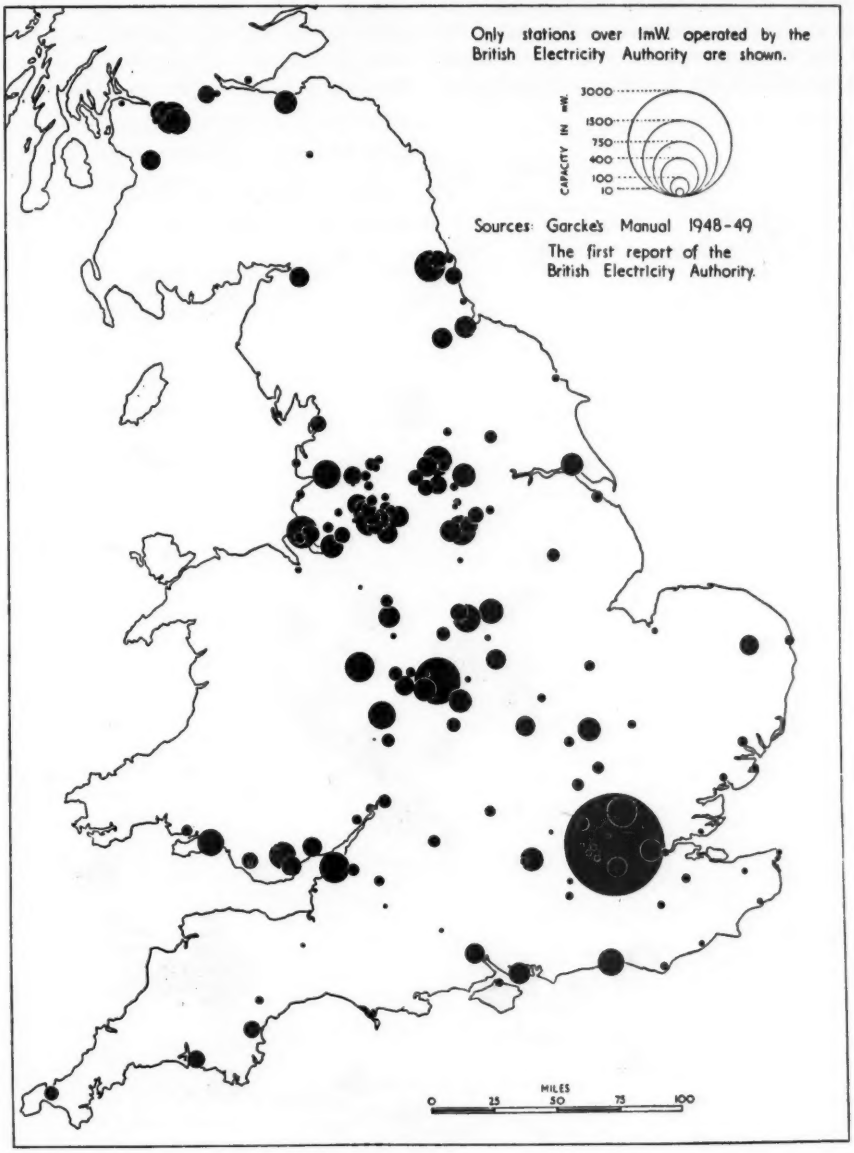


Fig. 4.—Coal-fired electricity generating stations. Capacity in 1948.

break in statistical continuity is unlikely therefore materially to affect the argument which follows, except that the relative importance of Trent-side stations is reduced when capacity and not output is considered.

Fig. 5 shows the distribution of new electricity generating capacity which the British Electricity Authority intends to establish between 1948 and 1960–65 approximately.[1]

The salient change in the economic

[1] No clear indication is available of the amount of pre-1948 plant which will still be operating in 1960–65, and therefore a composite map comprising old and new plant in 1960 cannot be given. Demand for electricity continues to increase so it may be assumed that much of the pre-1948 plant will still continue to function in 1960.

geography of electricity production in Britain is clearly apparent on this map. It is the line of large new stations in the East Midlands which are under construction or

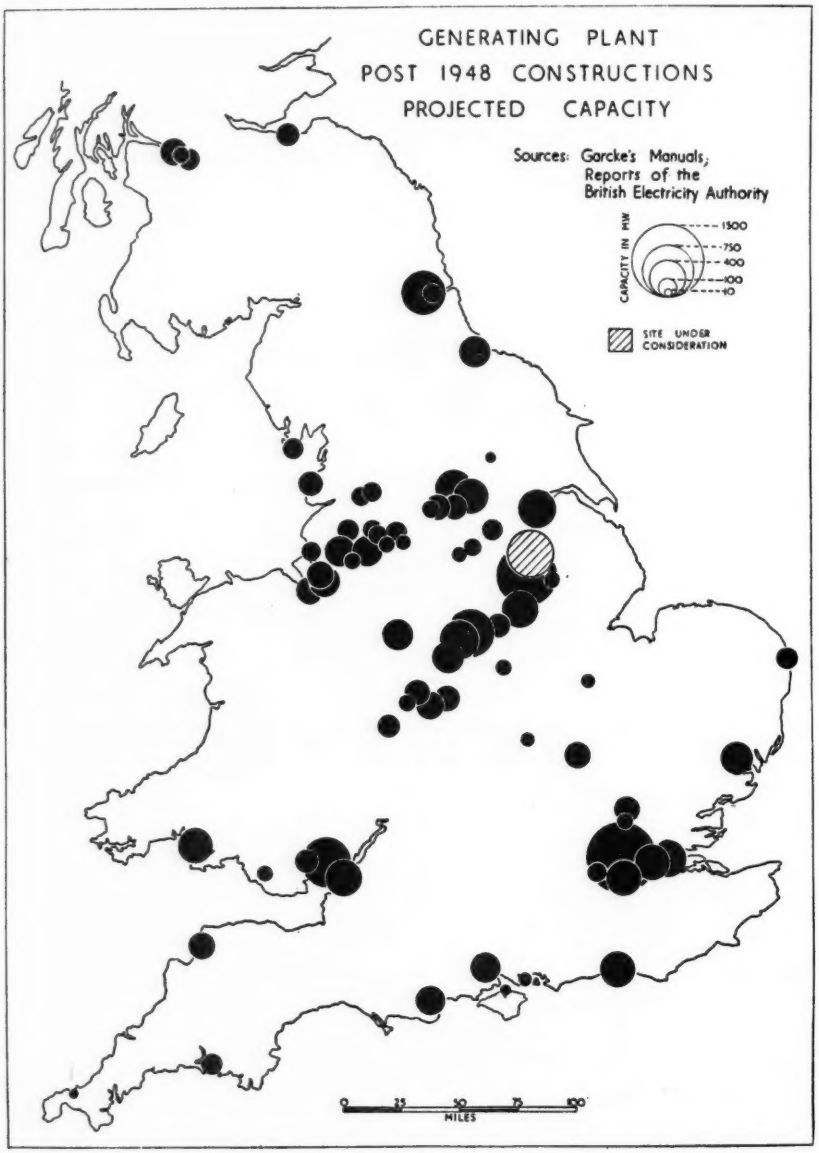


Fig. 5.

are planned for sites on the banks of the River Trent. The ultimate capacities of some of these stations are considerably understated on the map for the exact complement of generators to be installed after 1957 has not yet been officially published. (The exact site for the shaded station was still in doubt when this map was produced.) If the plan is fully implemented it is probable that this will be by 1965–70 the largest concentration of generating capacity and output on any equivalent length of non-estuarine river anywhere in the world. It will surpass both the International Rapids section of the St. Lawrence Seaway Project and the

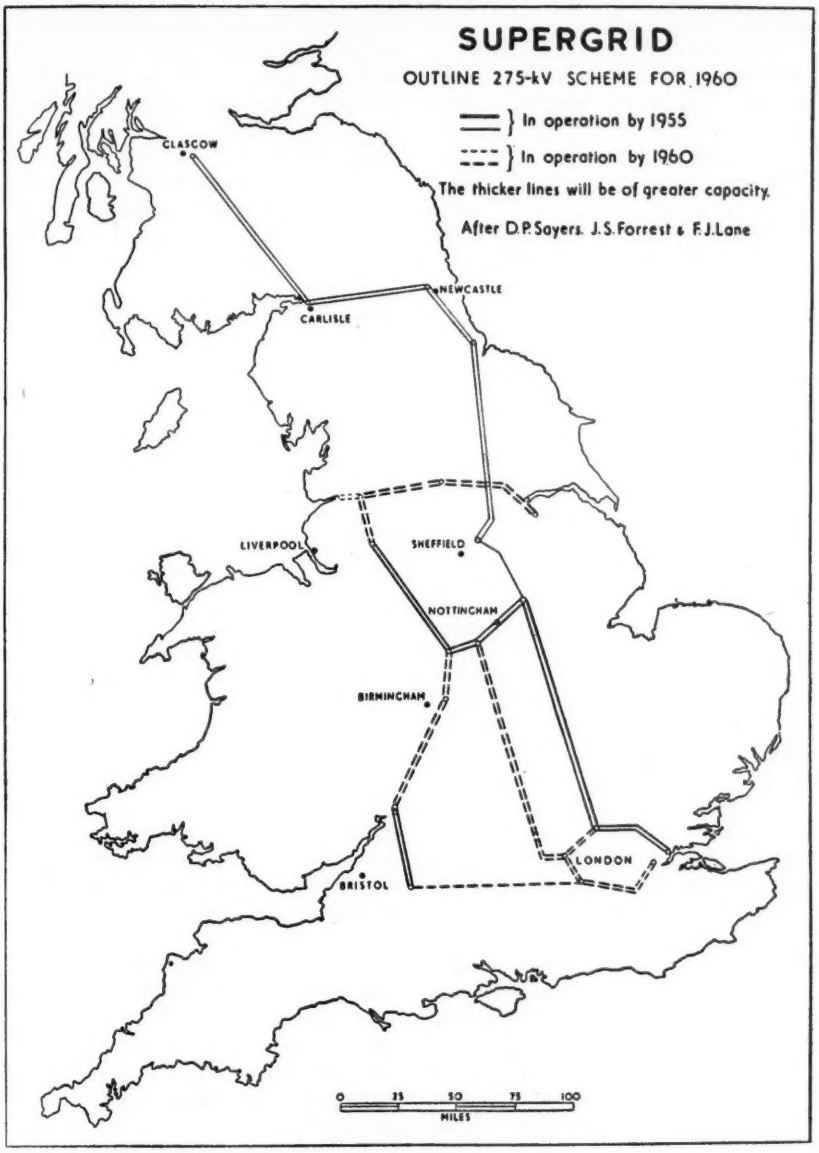


Fig. 6.

Tennessee Valley Authority in generating capacity and output.[1] Taking unofficial data into account the intended capacity on the river by 1965–70 should be not less than 4,700,000 kW and the annual output may well exceed 25,000,000,000 kWh.

[1] Doubts concerning the truth of this statement may be dispelled on reflection that Britain is a densely populated industrial country and that this line of new stations stands roughly at the centre of the great industrial belt.

Some of this plant is already in use to-day but the greater part of it and the change in the distribution pattern are yet to come. By 1960, however, a considerable change will probably have been wrought with the completion of the first stage of the new 275 kV. interconnection and transmission system (Fig. 6), much of which is to be constructed so that it may be modified later to operate on 380 kV. if required.

The main axis of this transmission

system, namely, the short S.W.–N.E. line running roughly from Birmingham through Nottingham to Newark, is closely related to the Trent-side power station developments. As the very large stations on the Lower Trent are constructed it will probably be extended to them and beyond into west Yorkshire thus forming a 'ring-main' in central Britain. From this 'ring-main' current may then readily be transmitted to other parts of the country by means of the other lines of the new Grid.[1]

The Reasons for the past Stability and present Changes in Pattern

This short account of the development of the pattern of electricity production in Britain poses two major questions. Why did the pattern remain constant from 1920 until after 1948, notwithstanding technical advances in electricity transmission? Why should the pattern be now in process of change?

Fundamentally there would appear to be a common answer to both questions, but as is often the case the simplicity of the fundamental is vitiated by a number of important but subsidiary considerations which blur its outlines.

Of the subsidiary considerations the less important was the organisation of the industry under a system of local monopoly. The power of local electricity authorities was, however, greatly weakened by the setting up after the 1926 Electricity Act of the Central Electricity Board with extensive powers over the industry. Indeed, plans for the locational changes now taking place were in part initiated by the Board.

Rather more important was the 'Grid System' itself and the effects it had. The 'Grid' was completed for most of the country by 1933. It was constructed to accommodate a pre-existing pattern of power stations which it therefore tended to perpetuate. Furthermore, it reduced the need for new plant for several years. By linking up all the major stations greater security was afforded to the supply of current. Thus a much smaller proportion of emergency plant was required than when each station produced electricity in isolation. There existed, therefore, a large pool of former spare plant which was brought into use as the demand increased during succeeding years. This was a large economy sufficient in itself to cover the cost of the 'Grid' and, by reducing the need for new stations, it may be deemed to have postponed the need for a serious review of locational policy. But even if this condition had not existed it is unlikely that any change in pattern would have been approved before the outbreak of war in 1939.

The Fundamental Reason

The fundamental reason for the constancy of the pattern between the wars, and for the process of change now occurring is concerned with two factors: firstly, the comparative cost of transmission and coal transport and, secondly, progress in plant design. The important points with regard to the former are listed below.

1. Electrical transmission cannot 'compete economically with transport of coal by sea where the rail haul from the pithead to port is short and the coal is delivered direct from sea-going colliers to the bunkers of stations situated near the load centres' (concentrated markets).[2] This condition is likely to remain effective. It is therefore uneconomic to transmit current in bulk on a planned schedule from either South Wales or Northumberland to London.

2. It is uneconomic to construct power lines solely for the purpose of transmitting mid-load or peak-load current from coal-field power station to distant market. A power line used unidirectionally for this purpose alone cannot compete with rail-borne coal carried to a power station near the market.

3. Transmission of base-load current in Britain is cheaper than rail transport of coal over distances greater than 50 miles and possibly somewhat less.

All these conditions have probably applied to a large degree for at least twenty years. Improvements in the technique of transmission have reduced its minimum competitive distance, but not greatly. Interior southern England, for

[1] The thinner lines on Fig. 6 are intended for interconnection only, not bulk transmission. A more up-to-date version of Fig. 6 is shown in the Sixth Report of the British Electricity Authority, 1953–54, Appendix 25.

[2] D. P. Sayers, J. S. Forrest and F. J. Lane, '275 kV Development on the British Grid System,' *Proceedings of the Institution of Electrical Engineers*, Vol. 99, 1952, Part 2, p. 585. Fig. 6 here is based on Fig. 6 in the above-mentioned article.

example, has long been well within the range of cheaper transmission from the East Midlands coalfields. Yet the electricity authorities did not begin to take advantage of the apparent economy of large-scale, bulk transmission of base-load current until very recently.

Progress in plant design is the chief reason for this. Throughout the inter-war period the efficiency of the turbo-generator was consistently being improved. Less and less coal was required to generate a given amount of current and each new station was appreciably more efficient than its predecessors. Coal (excluding transport costs) comprised a very large proportion of the works costs of production in Britain ; therefore economies achieved here were likely over a period of time far to outweigh economies in transport costs which amounted to a much smaller proportion of total works costs. Furthermore, economies in transport through transmission and a change in locational policy could have been made once only and thereafter would have yielded a constant return in proportion to the capacity of plant installed. Economies in fuel consumption were progressive and therefore tended to yield a greater proportional return as more and more new plant was commissioned. Thus it is possible to argue that attention was concentrated on the larger economy and the smaller was neglected.

But this is not the complete case, for the more extensive use of transmission from inland coalfield to distant market before the war would probably have yielded not a constant but a decreasing return over a period of time in proportion to the capacity of plant installed. This is because the reduction in coal consumption through improvements in design had a marked effect on the use to which individual generating sets were put. The newer sets, being less costly to run, were used most. They operated on base-load. The oldest sets producing current at high cost were used as peak-load plants. The rate of obsolescence (cf. the history of aircraft design) was so rapid that new sets did not remain on base-load for more than a few years. They were then relegated first to mid-load and later to peak-load working.[1]

Plant had to be designed and located to fulfil each of these purposes in turn during the course of its existence and since it spent only a small part of its life on base-load, which amounts to no more than one-third of the daily maximum demand, it was more economic for stations to be located near their intended markets than near the source of their coal.[2] For as we have already noted it was and is in the main uneconomic to transmit current other than base-load over any considerable distance.

This may be an overstatement of the case in favour of the former national electricity authorities for there are some in the industry who have long advocated large-scale transmission from the inland coalfields. They were unable formerly to press home their arguments because, although the short-term advantages were reasonably clear, the long-term advantages were not. And the progressive economies from the increasing efficiency in coal consumption made those from transmission appear negligible even if they were not in the long run negative.

The change in locational policy favouring transmission of base-load current from stations near the inland expanding coalfields has come about through the expectation that substantial improvement in the efficiency of coal consumption was almost at an end.[3] This made transport costs a relatively more attractive field for economy than had been the case as long

[1] This paragraph is a simplified version of what actually happened.

[2] It must be remembered, however, that coalfields and concentrated markets for current frequently coincide geographically in Britain. London is the great exception.

[3] V. A. Pask, *Modern Trends in the Design and Location of Electrical Generating Stations*, Fourth World Power Conference, London, 1950, Paper 8, p. 12. 'In the past, improvements in the heat cycle were rapid, and each new station tended to be appreciably more efficient than its predecessors ; hence it began life as a base-load station and, becoming obsolescent, ended as a peak-load station. The rapid approach to the limits of the heat cycle makes it clear that in future the cost of production will depend more upon the location of the generating station than upon its date of construction.

'The best sites will therefore be occupied by stations designed to carry base-load to the end of their lives. Such a course will release the designer from the obligation to make the plant suitable for intermittent operation in its later years, and this will assist in the simplification of the design. The corollary to these base-load stations will probably be two other groups of stations on what might be called " secondary sites," one group designed for mid-load operation, and the other for peak-load operation. These stations will be essentially of lower capital cost than the base-load stations.'

as technical progress continued in plant design. Thus the expectation that the maximum efficiency at the generating station was shortly to be reached compelled the electricity authorities to take full account of both major geographical components, the location of the market and the location of appropriate coal production, in the cost structure of the industry. No longer could they consider the one and neglect the other as was formerly the case.

As a result of this expectation, instead of almost complete market location of electricity production, a geographical and technical reorganisation is now taking place on the following lines :

1. Large new base-load stations to operate as such virtually to the end of their lives, or at least for a much longer period than before the war, are being established on the banks of the Trent and to a smaller extent the Yorkshire Calder and Aire. All these stations will be near the low-cost, expanding, inland coalfields which must, for reasons peculiar to the geography of coal production, provide in future an increasing proportion of the coal needed for electricity production. Sites on the rivers will provide the large volume of cooling water required. Electricity will be exported from here to the outer parts of London, interior southern England and south Lancashire. Northumberland is unlikely to produce sufficient coal of the appropriate kind for greatly increased use to be made of cheap coastal transport. South Wales coal is both higher priced and much too valuable for other purposes to be considered ; nor is the cost advantage of coastal transport so great. This impending shortage of coal available for coastal shipment means that a larger area of southern England must be dependent for a larger proportion of its supplies on transmission from the Trent than might otherwise have been necessary.

2. Base-load stations will still be constructed to supply local markets in all those many areas of the country where either coal production and market for electricity coincide or where the market is near navigable water by which adequate amounts of sea-borne coal can be received direct from coastal coalfields. South Lancashire, it should be noted, although possessing its own coalfield must import some base-load current from Yorkshire and the Midlands because its coal production is and will remain inadequate to satisfy local demands. Already only 60 per cent. of the coal used in Lancashire power stations comes from local mines.

3. Peak-load and mid-load stations, specially designed to operate as such throughout their lives, are likely to be built in all the main consuming areas. Transport costs do not favour extensive transmission for these types of stations.

That these changes should occur now is clearly because it was anticipated that fundamental technical-economic progress had virtually ceased. Accordingly the spatial-economic factor in the cost structure of the industry was up-graded in importance and those who control the industry had to take fuller account of the cost factor derived from the relevant circumstances of economic geography. As a result the range of economically and technically possible locations for specific operational types of plant (i.e. base-load, mid-load or peak-load) became, in plans for the future, far more stringently limited than formerly.

The reasons for this greater restriction might be summarised in the phrase ' perfect technique demands perfect location.' Surely this is a dictum, albeit perhaps a neglected truism, with considerable relevance to the realistic study of industrial location in general.

In the event, however, the expectation of the early attainment of perfect technique has been proved somewhat premature. Efficiencies of the order of 36 per cent. are now envisaged as compared with the limit of around 30 per cent. anticipated as recently as 1950. Such economies in coal consumption make savings in transport costs appear puny by comparison. For the present, therefore, the economics of the new locational policy must depend on the rapidly increasing demand for coal by the generating industry and the changing pattern of coal supply,[1] both considered in relation to the comparative cost of transmission and coal transport by rail and to the ever-increasing demand for current.

[1] The inland coalfields of the East Midlands and Yorkshire must provide the greater part of the additional coal required.

MILK LIBERATION

At the Oxford meeting of the Association, the physiology of milk liberation was discussed at a symposium which took place on the morning of September 7, 1954, between Sections I (Physiology) and M (Agriculture). Papers were contributed by Professor H. D. Kay, F.R.S. ('Introductory and general background'), by Dr. B. A. Cross ('Physiological mechanisms in suckling and milking'), by Mr. A. S. Foot ('Applications of recent findings to the milking cow') and by Dr. Mavis Gunther ('Clinical aspects of the milk propulsion reflex'). The last paper was accompanied by a film showing certain lactation phenomena in the human subject, and the other papers were illustrated by lantern slides.

The following digest of the symposium has been prepared by Professor Kay.

Introductory and General Background

The phenomena associated with the formation of milk and its discharge from the mammary gland have received increasing attention from physiologists and biochemists in recent years, and part of the new knowledge derived from their research has been applied both on the dairy farm and in the clinic. The term 'liberation' was used by those arranging the symposium to limit the latter to a portion only of the large field of lactational physiology.

The 'liberation' of milk, a process which, with minor modifications, is very similar in character in all mammals from the mouse to the whale, consists of three operations; first, the discharge of milk from the cells which actually form it—the cells lining the alveoli of the mammary gland—into the alveolar spaces; second, the pushing through of the secreted milk from the alveolar spaces into the larger ducts and milk sinuses without, however, any milk leaving the gland (a phenomenon known to the dairy farmer as 'let-down' and called by some physiologists 'ejection' and by others, perhaps more accurately, 'pertrusion' a pushing through of milk); third, the milking or suckling process by which milk is withdrawn from the gland by a suction on the gland orifice or a pressure from without on to the exterior surface of the teat sufficient to overcome the resistance of the teat sphincter present in many species or the analogous apparatus in certain other mammals.

As regards the first operation, the rhythmic cycle of milk formation, accumulation and discharge from the milk-secreting cell lining the alveoli, it seems certain, from histological and other evidence, that these cells preserve their integrity during a large part of a lactation and do not, as was once thought, undergo complete breakdown and re-formation with each discharge of synthesised milk. How many times during 24 hours an average active cell of a gland in full lactation goes through this cycle of milk formation and discharge we do not as yet know. During this first stage, the intra-mammary pressure on the secreted milk is very slowly rising.

The second stage of milk liberation does not normally take place until shortly after actual suckling or milking begins. After an induction period between the beginning of suckling or milking and the active flow of milk, a period which in the cow is from 40 seconds to about a minute, there is a dramatic rise in intra-mammary pressure to double or more its value before milking. This rise in pressure is now known to be due to a neurohormonal reflex, the tactile stimulus of the sensory nerve endings in the teat or nipple or in the gland itself, caused by suckling or milking, bringing about the reflex liberation of the hormone oxytocin, or most probably both oxytocin and the vasopressor hormone, vasopressin, from the posterior-pituitary gland. These

hormones, both of which are known to be polypeptides and which have recently been synthesised in the laboratory, are discharged into the blood stream and eventually (in less than a minute in most mammals investigated) reach the mammary gland. Oxytocin causes certain structures in the mammary gland—the so-called myoepithelial cells, cells shaped rather like tiny starfish with their arms surrounding the alveoli and the smaller ducts—to contract and effect the squeezing-out or pertrusion of the secreted milk from the alveolar spaces into the collecting ducts and, in those species where sinuses exist, into the sinuses or cisterns of the gland. Oxytocin is active in astonishingly high dilution in bringing about myoepithelial contraction. Pure vasopressin seems also to be active—but with only a small fraction of the activity of pure oxytocin.

The let-down, ejection or pertrusion phenomenon in so far as the intra-mammary pressure is concerned, does not normally persist for more than a few minutes, even if the gland is not milked, and there is evidence that the circulating oxytocin is either destroyed or taken out of the circulation fairly rapidly.

Physiological Mechanisms in Suckling and Milking

Much detailed information on the second stage of milk liberation—the let-down stage—has been derived from recent experiments in which the lactating rabbit has been used. This animal, whose lactational mechanisms seem to bear a close resemblance to those of the cow or the human subject, makes an almost ideal experimental animal for work on the role of the central nervous system and the endocrine glands in lactation. Rabbits are extraordinarily capable milk producers, peak yields of milk which contains over 30 per cent. of solids (cow's milk contains some 13 per cent. of solids) may approximate to one-tenth of the body weight daily. These animals exhibit a well-defined let-down during suckling, which may be rendered visible by cannulating one of the unoccupied mammary glands. The rise in intra-mammary pressure may be sufficient to maintain a vertical column of milk 15 cm. high in the cannula. Anaesthesia blocks this let-down or pertrusion reflex.

Myoepithelial cells are very abundant in the rabbit mammary gland. It has been shown that the myoepithelium responds to a variety of drugs—acetylcholine, pilocarpine, barium chloride, histamine, as well as to the posterior pituitary hormones. It also responds to direct electrical or mechanical stimuli. It is rather surprising, therefore, that so far there has been no evidence for the existence of a direct nervous control of let-down.

The effect of posterior pituitary extracts can conveniently be studied by recording on a moving drum the milk ejection responses from cannulated mammary glands in the anaesthetised rabbit. In these animals the threshold dose of whole posterior pituitary extract is low—of the order of 1–2 milli-units. Maximal effects are obtained with about fifty times these doses. With the larger doses, the level of milk in the cannula rises rapidly after about 10 seconds, reaches a peak in about 60 seconds and then slowly subsides. After such injections, all the mammary glands of the rabbit are affected more or less simultaneously, but the effect of a mechanical stimulus is localised to the stimulated lobule, the response occurring after one second only and usually subsiding within one minute. The buffeting of the mammary gland by the young suckling can, at best, account for only a very small proportion of the milk removed by the young.

It is necessary to establish not only that posterior-pituitary extracts can stimulate *natural* let-down, but that the intact posterior pituitary gland is capable of doing so. It is also necessary to show that removal or destruction of this gland abolishes the capacity for natural let-down. Experiments in the rabbit involving direct electrical stimulation of the nerve supply of the posterior pituitary demonstrate that milk ejection closely duplicating the responses given by appropriate doses of posterior pituitary extract (approximately) equal to the response given by 50 milli-units of a whole posterior pituitary extract) can be obtained by such stimulation. Litters can withdraw normal amounts of milk from anaesthetised

does after such stimulation. Milk let-down has resulted from electrical stimulation of any part of the entire neurohypophysis and its nerve supply. If, however, electrolytic lesions are placed so as to interrupt the nerve fibres to the posterior pituitary, although the animals show no outward signs of the effect of such lesions and their nursing behaviour is apparently normal, let-down does not take place nor are the sucklings able to obtain more than a fraction of the full milk yield unless posterior pituitary replacement therapy is given.

As regards the identity of the hormone normally responsible for the let-down (= milk ejection or milk pertrusion) reflex, samples of the highly purified oxytocic and vasopressor-antidiuretic polypeptides prepared by du Vigneaud and his colleagues have been compared. It has been shown that oxytocin is approximately six times as active in milk ejection potency, weight for weight, as the antidiuretic hormone. It has, in fact, been demonstrated in lactating rabbits that a diminution in urine flow does occur in most cases after suckling. But the amount of the antidiuretic hormone secreted is so small that it is well below its threshold value for milk ejection, and this hormone can be ruled out as the physiological excitant of the let-down of milk in the rabbit. Oxytocin appears to be, therefore, the milk ejection hormone. From a comparison of the let-down responses recorded from a cannulated teat during suckling with the responses to different doses of oxytocin injected intravenously, and also from the results of another method depending on replacement therapy, it is concluded that, in the rabbit, amounts of the order of 50 milli-units of oxytocin are released into the blood stream in response to the natural stimulus of suckling.

As in the cow and still more obviously in the lactating woman, so also in the rabbit, emotional disturbances at suckling time lead to diminution or to complete inhibition of the milk ejection reflex. It has been shown in the rabbit that the disturbance of the reflex involves a failure to release oxytocin, though in a few of the cases examined there is also a peripheral effect, as shown by a reduced milk ejection response to 50 milli-units of injected oxytocin. Anaesthesia restores the response to this dosage of oxytocin.

Stimulation of the sympathetic centres in the hypothalamus inhibits the milk ejection response in a similar way to the injection of 1–5 micrograms of adrenalin. Further experiments have demonstrated a dual inhibitory pathway from the hypothalamus—a direct nervous pathway to the mammary gland, responsible for the immediate inhibitory effect and a hormonal pathway via the adrenal medulla. The latter inhibitory effect is probably mediated by the pronounced mammary vasoconstriction caused by adrenalin which prevents the access of circulating oxytocin to the myoepithelial tissue. The sympathetico-adrenal discharge certainly interferes with milk removal by the suckling young. It is, however, the *central* inhibition of oxytocin release which takes precedence in emotional disturbance of let-down in the rabbit.

Application of Recent Findings to the Milking Cow

How far can these recent findings be applied on the dairy farm? Most of the dairy cows in Britain are now milked by machine, and it is in connection with machine milking that the physiological findings are of immediate importance. The main practical concern is with the second and third phase of milk liberation —with the pertrusion or let-down phase and with the physical removal of milk subsequently.

As regards pertrusion, the characteristics of particular interest in machine milking are (1) the transitory nature of the let-down reflex, whereby rate and duration of milking assume great importance, (2) the character and perhaps the intensity of the stimulus that produces the neurohormonal reflex, since these may affect the amount and quality of milk produced at a milking or series of milkings, (3) the possible inhibitory effects of methods of handling or mishandling the cow during milking.

As regards the physical removal of milk past the teat orifice, the essential action of the milking machine is to reduce the air pressure on the outside of the teat to a sufficient extent to open the teat orifice and allow the flow of milk. To avoid

congestion of the teat tissues with blood, the suction is applied rhythmically, pulsations usually being from 30 to 60 per minute.

Whether hand or machine milking is used, great variations in the *rate* of milking from cow to cow are well known to exist. These are not only important from the standpoint of labour economy, but recent work at Shinfield has shown that *rate* of milking—which is surprisingly constant for a given cow under standard machine-milking conditions—is related both to milk yield and to susceptibility to mastitis infection. In general, the rapidly milking cow gives a greater lactation yield but is rather more prone to mastitis than the slow milker. A slow rate of milking is, however, almost entirely an anatomically conditioned effect and not associated with a possible defect of the let-down reflex. With a cannula of suitable size in the teat canal, the slow milker becomes a rapid milker. (The shape of the milk-flow curve (rate/time) of a given cow is also remarkably constant under standard milking conditions even in different lactations.)

Nevertheless, variations in the physiological efficiency of the let-down reflex may affect the quantity of milk obtainable at a milking. Is this quantity, in the cow, dependent on the actual amount of oxytocin released by the posterior pituitary? This appears to be the case in the sow, where yield of milk at a single hand milking is related to the amount of oxytocin injected. So far there is no good evidence that the amount of oxytocin naturally released in the cow changes very much from one milking to another. It is more important from the practical point of view to pay main attention to the time relationships between the stimulus to let-down and the attachment of the milking machine. Thus, the teat cups should be attached not more than three minutes after the effective stimulus to let-down (usually the washing of the udder) has been applied. This stimulus may be conditioned (in the Pavlov sense) to be either the washing of the udder, the feeding of concentrates, the noise of the milking machine or certain other sensory stimuli which become associated, in the cow's mind, with milking—hence the need for keeping to a regular routine at milking time so that effective conditioning may be maintained. There is evidence that in the slow-milking cow the let-down reflex fades before mechanical removal of the milk has reached the stage of completion achieved in the fast milker.

The milking machine itself can and does influence milking rate. The extent of 'vacuum' employed, the pulsation rate, the design of the teat cup, are all known to play a part. An increase in the 'vacuum' from the normal of about 11″ Hg to, say, 20″ Hg will substantially increase the peak rate of milk flow, but may suck too much of the teat into the cup and thus block the removal of 'strippings'. A higher 'vacuum' may also damage the udder tissue and increase the risk of mastitis. A rise in pulsation rate also increases to a limited extent the rate of flow.

In attacking the practical problems involved in improving a milking system on a dairy farm, the removal of slow-milking cows from the herd is one of the most obvious steps. Either regular cannulation of the teats or an operation to increase the effective size of the teat orifices are alike impracticable. Slow-milking animals are unlikely to improve in later lactations; the character seems to be fairly highly heritable and such cows certainly upset milking routine with waste of time and labour.

There is much to be said for bail or parlour milking, in that stimuli to let down or pertrude milk can more easily be arranged to occur at about the right time, i.e. shortly before the teat cups are applied. Udder washing should be carried out as near to milking time as possible. The 'vacuum cleaning' of the udder with extensive hand or machine stripping is unnecessary; fairly large quantities of milk (some 20 per cent. of that originally present before milking) and even larger quantities of milk-fat are in any case left in the udder after a milking. Nevertheless, it is important to determine the optimum point at which the teat cups should be removed, neither too early (which usually leads to a low fat percentage in the milk) nor too late (which may result in increased incidence of mastitis). It is in the solution of these day-to-day problems that many modern dairy farmers combine both the results of scientific research and the accumulated experience of their ancient craft.

Clinical Aspects of the Milk Propulsion Reflex

The knowledge obtained from various other mammals in recent years on the let-down or pertrusion reflex (called by some, in regard to the human subject, the ' propulsion ' or the ' draught ' reflex) has important applications to human lactation and its management. A rather large proportion of human mothers have difficulties in feeding their infants from the breast, difficulties arising in part because emotional and psychological conditions control this function in the woman to a far greater extent than in other mammals. One result is that the draught reflex is fast becoming a tool for the study of the physiology of emotion.

In most women the draught is a consciously-felt sensation not easily described in words. As one might expect, it arises just before, or within a very short time after, the baby is put to the breast. When it arises before suckling begins, it is usually associated with thoughts relating to the feeding of the baby. The draught *sensation* fades more slowly than it appears ; it is sometimes painful in early lactation ; it is felt in the unoccupied, as well as in the occupied breast ; it lasts from a few seconds to about 1½ minutes, the sensation probably only occurring at the height of the reflex increase of intra-mammary pressure. Milk may at this stage drip from the unoccupied nipple. The draught *reflex*, i.e. the let-down reflex, occurs in almost all lactating women at almost all stages of lactation although the intensity appears to be only sufficient for the draught to be *felt* when the reflex has been reinforced by use or by conditioning.

The draught is not usually felt when the breasts are empty, but from the ready firing off of the reflex when the breasts are full, it appears that the actual distension becomes a contributory factor to any stimulus which is effective. When the breasts are full, the reflex may be triggered off by trivial, and by less and less specific, stimuli. There may perhaps be at this time either an optimal stretching and increased sensitiveness of the myoepithelial structure, or a heightened preparedness of the pituitary source of oxytocin as a result of the arrival in the central nervous system of sensory stimuli from the distended breast.

Clinically, early vigorous draught action, frequently indicated by leakage of milk from the free breast, is a good prognostic sign for a successful lactation, and it may well be that some part of the mechanism of the draught contributes to the maintenance of lactation. There is reason to believe that an effective let-down in the cow may also be associated with a persistent lactation, sensory stimuli from the mammary gland helping in the production, by the anterior pituitary gland, of prolactin, the hormone responsible for maintenance of lactation.

What happens inside a breast when the draught occurs and no milk escapes ? In the cow's udder, there are extensive milk cisterns to hold pertruded milk, but in the human breast there is no cistern, and the anatomical arrangement of the ducts beneath the areola where pertruded milk is presumably held is not only very different from one individual to another, but appears to possess rather limited storage capacity. The ducts are thin and distensible, sometimes anastomosing and sometimes not. When milk is retained in the breast after a draught reflex has taken place, the ducts probably distend along their length rather like the sac of a fountain pen than in large local bulges.

There is some experimental evidence in the human subject that the sphincter action of the nipple can be relaxed by warmth. If the nipple of the unoccupied breast is warmed and light suction is applied, a rapid flow of milk may follow from that breast when draught is established by putting the baby to the other breast.

It has been observed clinically that a febrile reaction, not infective in origin, may occur about the time of the inception of acute intra-mammary mastitis. This reaction may result from the increased intra-mammary pressure at a draught period forcing the escape of milk through the damaged mammary tissue into the circulation. The draught, and relaxation after the draught, are also important clinically if acute mastitis has turned to abscess formation. In such a condition of the breast, it seems reasonable to suppose that the infected material may, in some cases, be driven down into a duct and when the reflex is spent, this material

may, if not completely drawn off during the draught, run back and, as it were, shunt into another siding, thus spreading the infection.

The main characteristic of the draught, from the standpoint of ensuring effective lactation, is undoubtedly its sensitiveness to the emotional condition of the subject. Here an experienced and sympathetic attendant to the newly-delivered woman (who is often in an abnormal psychological state where apprehension is liable to flood in) may enable a normal draught reflex to become established and a satisfactory lactation to ensure, in a case which might otherwise founder on the rocks of misunderstanding or fear.

A lively discussion followed the symposium.

HEAT IN THE EARTH[1]

by

D. W. ALLAN

Endeavour Prize Essay

1. Introduction

The Earth's deep interior is unquestionably hot, but there is no other fact about its thermal state and history which has been decided with certainty, even though many eminent scientists have studied the matter. Fourier, who laid the mathematical basis for the study of heat conduction, gave a brief discussion of the cooling of the Earth (1820) in which he began the dubious practice of using the solution for a semi-infinite plane as an approximation. He pointed out that if the initial temperature, assumed constant, and the present surface gradient were known, then one could estimate an age for the Earth. Lord Kelvin (1862) later derived the same result, and thought he had sufficiently good numerical data to set a fairly precise upper limit for the age of the Earth, viz., 100 million years. This result made most geologists thoroughly unhappy, and they were still less pleased when he later revised his estimate to 25 million years. It was unfortunate that Lord Kelvin attached such weight to these estimates, for his reputation was such that this low figure became accepted as physics' final answer to the problem, and when the discovery of natural radioactivity at the turn of the century showed Kelvin's result to be inapplicable, geologists were encouraged in a mistrust of the mathematical-physical approach to theories on the Earth which has persisted to this day.

The existence of a source of generation of heat changed the problem of heat conduction in the Earth completely. Since the concentrations of radioactive substances in rocks were for a long time scarcely determined at all (even by 1940 most of the results were unreliable) most of the early discussions of the thermal history of the Earth which take into account the presence of radioactivity are of interest only for the type and behaviour of the analytic solutions used, the numerical values obtained being almost valueless. The most important accounts were those of Jeffreys (1929), which he later modified considerably (1952) because of errors introduced by the use of the semi-infinite plane solution, but which has been much quoted and used; two theoretical discussions by Lowan (1933, 1935) which gave solutions for an exponential distribution of radioactivity in a sphere; papers by Spicer (1937) and Van Ostrand (1940) containing tables of numerical solutions, and a very thorough discussion by Slichter (1941).

What was probably the most significant advance in the study of thermal conduction in the Earth since the discovery of radioactivity was made by W. D. Urry (1949) in a paper in which he calculated the surface heat flow for the Earth, using a solution due to Comenetz which took account of the fact that the production of heat by the radioactive elements has varied with time. It had not previously been appreciated what a great difference in heat production would occur in a time-span as long as that since the solidification of the Earth.

So far I have discussed the approaches to the problem of the Earth's thermal history from the standpoint of conduction theory. There is another entirely different approach to the whole question, that which assumes the presence of slow convection currents in the material of the Earth's

[1] Essay awarded the first prize in the 1954 competition sponsored jointly by the publishers of *Endeavour* and by the British Association. Mr. Allan is a Canadian and is doing post-graduate research in geophysics at the University of Toronto.

mantle. Much of a qualitative nature has been written on this theory by Pekeris, Vening Meinesz, and others, but it seems unlikely that any fairly precise conclusions as to the numerical values of the temperature and heat flow in the Earth's interior can be given. Urey in his book *The Planets* (1952) made rough calculations based on the convection hypothesis and on a suggestion he had made that the dense inner core of the Earth has been, and still is being formed by a process of slow differentiation, rather than having been formed rapidly at the beginning of the Earth's history, but he was able to come to no very definite conclusions.

I would conclude from this survey of the principal difficulties and achievements in the study of the Earth's thermal history that one cannot hope to give a complete and final treatment of the subject. The best that can be done is to examine some one hypothesis as to the mechanism by which the Earth is cooling or heating, and see whether it produces results which agree with observable facts. Since the model of an Earth cooling by conduction is the one most susceptible to exact mathematical treatment, as well as for other reasons discussed in the next section, I have chosen it as the subject of this essay.

2. Observational Evidence Bearing on the Problem

The knowledge of the present state and past history of the Earth which science at present possesses is imperfect, but sufficient to justify attacking the problem of its thermal history.

From astronomy and geodesy it is known that the Earth is very nearly a sphere, and the seismic study of earthquakes has shown that, apart from thin surface layers, the material of the Earth is distributed in a spherically symmetric manner. Furthermore, this same seismic work shows that, so far as the transmission of elastic waves is concerned, the Earth behaves as a solid to a depth of nearly one-half its radius, and may contain further solid material at still greater depths. This does not prove that the material of the mantle cannot flow over long periods, but does show that the assumption of a mantle which behaves as a rigid solid is not an unreasonable one. The fact that A. E. Scheidegger and J. T. Wilson in a series of recent papers have been able to explain many geotectonic features on the basis of the thermal contraction of a solid Earth adds further strength to the assumption. The effect of the liquid part of the core will be to equalise the temperature in the liquid (more correctly, it causes the temperature gradient to become approximately the adiabatic) and the error produced by treating the entire Earth as solid will be negligible in the greater part of the mantle.

Another argument for treating the Earth as a radially symmetric solid sphere in the discussion of its thermal history is that the measurements of heat flow from the Earth's interior which have been made at widely separated places on the Earth are quite consistent, and even the measurements of heat flow through the ocean floor do not differ significantly from those on land. This, incidentally, provides another reason for thinking that the mantle is rigid, since steady convection currents in its material should be reflected in greatly different surface heat flows. These conclusions are not at all certain, however, for only about fifty determinations of heat flow have been made, and in only some seven regions. There are not enough results to justify any elaborate statistical analysis, but a comparison of several of the groups in pairs shows that the means for Canada and South Africa are not significantly different, and that the same applies for the British and Atlantic results. On the whole, therefore, there is no reason to believe that the flow of heat from the Earth's interior is not practically constant over the Earth's surface, including the oceanic regions, with a value of about $1{\cdot}2 \times 10^{-6}$ cals./cm.^{2}sec. This is equivalent to 38 cals./cm.^{2}cm. year, which means a total loss from the Earth's surface of about 2×10^{20} cals./year. The rate of heat loss by vulcanism is probably less than 1/100 of this, according to the estimates of Lotze (1928) and Verhoogen (1946), and even under the most extreme hypotheses it would be an order of magnitude less (Birch, 1951).

The surface temperature of the Earth is essentially determined, not by the amount of heat arriving from the interior, but by the balance between heat from the sun

and that radiated back out into space. Furthermore, the effect of diurnal and annual variations, and even of climatic changes such as the ice ages, is unimportant when considered on a time scale of billions of years. The average surface temperature changes by about 40° C. from the polar to equatorial regions, but this difference may be neglected in a study of the interior temperatures. The present surface temperature would be greatly different from that in the past only if there had been extensive changes in the atmosphere, if the rate of energy radiation by the sun had changed considerably, or if the Earth had been at a different distance from the sun. I have made calculations of the possible extent of these effects, and find that the most extreme variations which are at all plausible are only about $\pm$ 60° C. Since the temperature in the oceans below two kilometres, which may be expected to be a good indication of the average temperature over the Earth's entire surface, is uniformly around 2° C., the surface temperature can be assumed to have been 0° C. throughout its history with the assurance that this will not be wrong by more than a few tens of degrees.

The remaining matters which determine the thermal conduction problem are the values of the thermal conductivity and thermal diffusivity, and the distribution of radioactivity within the Earth. The type of radioactive distribution I have used is discussed in the next section, and numerical values therefor are given in Section 4. W. D. Urry (1947) and, more recently, R. J. Uffen (1952) have examined the variation with depth of the conductivity and diffusivity, and their results show that, although the assumption that these are constant in the mantle is not a good one, it should not cause serious errors, at least for the upper part of the mantle. Furthermore, it can be shown that no solution reasonably amenable to numerical calculation can be obtained unless these are assumed constant. However, the variation in the conductivity can be taken into account partially in calculating the heat flow. The value I have used for the diffusivity is ·007 cm.²/sec., and I have employed Uffen's (1952) values for the conductivity.

Finally, although the "age of the Earth," that is, the time since its solidification, can in theory be determined from the observed surface heat flow, it turns out in practice that this heat flow is very greatly dependent on the near-surface distribution of radioactivity, which is fairly uncertain. The best procedure, therefore, is to use an age for the Earth derived by other means, and then use the observed heat flow only as a check on the radioactive distribution adopted. Fortunately, the age of the Earth is now fairly well known. One estimate, based on the recession of the moon caused by tidal friction, gives an upper limit of 4×10^9 years for the Earth's age (Jeffreys, 1952). I have examined possible modifications of this result, for example by the theory of continental growth, and find that the value is not likely to be lower than this, and may well be larger. Another method, based on the apparent recession of the nebulae, gives an age for the Universe as a whole (and therefore an upper limit for the Earth's age). Until recently the value thought correct was 2×10^9 years, but in 1953 Dr. W. Baade discovered that the distances of all the nebulae had been underestimated by a factor of nearly three, and the value now accepted is $5 \pm 1 \times 10^9$ years. The age of the oldest known rocks, about 3×10^9 years at present, provides a lower limit for the age of the Earth. The most important method, so far as thermal work is concerned, is that based on the variations observed in the isotopic abundances of lead in galenas, since this yields the time since the solidification of the upper part of the Earth's mantle. The value formerly accepted was about $3{\cdot}5 \times 10^9$ years, but recent investigations using larger numbers of samples by R. D. Russell (1954) and by me (to be published) indicate that the correct value is higher, and perhaps as high as $4{\cdot}5 \times 10^9$ years. In the present calculations I have used an age of 4×10^9 years, and from what is said above it is clear that this should not be in error by more than $\cdot 5 \times 10^9$ years.

3. The Mathematical Solution

The equation of heat conduction for a radioactive, spherically symmetric Earth is

$$\rho c \frac{\partial T}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r}\left(K r^2 \frac{\partial T}{\partial r}\right) + H(r, t) \quad (1)$$

where T stands for the temperature,

$H(r, t)$ is the rate of production of heat per unit time and volume

c is the specific heat, and K the thermal conductivity.

ρ represents density. (The thermal diffusivity, $k, = K/\rho c$)

The type of radioactive distribution which will be used is one in which the radioactive substances are distributed in shells of constant concentration, for this can be made to fit any given distribution fairly well by taking a sufficient number of shells, and gives rise to a solution well suited to calculation with different sets of values for the concentrations.

The rate of production of heat due to radioactivity, H (r, t) may be written

$$H(r, t) = \sum_i h_i(r)e^{-\lambda_i t} \qquad (2)$$

where $h_i(r)$ is the heat production per unit volume and λ_i the decay constant of the ith radioactive substance. The time t is measured from the time of solidification t_0 years ago. Equation (1) may then be written

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial K}{\partial r} \cdot \frac{\partial T}{\partial r} + K\left(\frac{2}{r}\frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial r^2}\right) + \sum_i h_i(r)e^{-\lambda_i t} \qquad (3)$$

The boundary conditions for the temperature $T = T(r, t)$ are

$$T(a, t) = 0 \qquad (4)$$

where a is the radius of the Earth, and

$$T(r, 0) = f(r) \qquad (5)$$

where $f(r)$ is the initial temperature distribution.

If $T = T' + T''$, where T' satisfies the equation

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial K}{\partial r} \cdot \frac{\partial T}{\partial r} + K\left(\frac{2}{r}\frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial r^2}\right) \qquad (6)$$

with boundary condition (4) and

$$T'(r, 0) = f(r) \qquad (7)$$

then T'' must satisfy (3) with the boundary condition (4) and

$$T''(r, 0) = 0 \qquad (8)$$

Thus the solution for a radioactive Earth with a given initial temperature distribution is the sum of the solutions for a non-radioactive Earth cooling from the given initial temperature distribution and for a radioactive Earth heating up from zero temperature. The solution to the first problem is well known and will be given later (in section 4). Consider the latter problem, and drop the superscript $''$ from T''. It is easy to show that the effect of each radioactive substance may be evaluated separately and that their sum will give the total effect. Thus in discussing the solution, $\sum_i h_i(r)e^{-\lambda_i t}$ will be replaced by $h(r)e^{-\lambda t}$.

Assuming that the radioactive substances are distributed in shells in each of which their concentrations are constant, write

$$h(r) = \begin{matrix} H_0 \ (0 \le r \le r_1) \\ H_1 \ (r_1 \le r \le r_2) \\ \cdots\cdots\cdots \\ H_n \ (r_n \le r \le a) \end{matrix} \qquad (9)$$

and

$$\begin{matrix} Q_0 = H_0 \\ Q_1 = H_1 - H_0 \\ \cdots\cdots\cdots \\ Q_n = H_n - H_{n-1} \end{matrix} \qquad (10)$$

If $T_m(r, t)$ is the solution of

$$\rho c \frac{\partial T_m}{\partial t} = \frac{\partial K}{\partial r} \cdot \frac{\partial T_m}{\partial r} + K\left(\frac{2}{r}\frac{\partial T_m}{\partial r} + \frac{\partial^2 T_m}{\partial r^2}\right) + \begin{cases} Q_m e^{-\lambda t} & (r_m \le r \le a) \\ 0 & (0 \le r \le r_m) \end{cases} \qquad (11)$$

then it is easy to show that the temperature T_j in the jth shell is given by

$$T_j = \sum_{m=0}^{j} T_m \qquad (12)$$

i.e. the general solution can be built up from a number of solutions of one type. This cannot be done if K and ρc are not continuous across the shell boundaries, i.e. the general solution for the temperature cannot be found by the superposition of a number of simple solutions unless K is treated as a constant. With this simplification, the general solution can be built up from the solution of the equation

$$\rho c \frac{\partial T}{\partial t} = K\left(\frac{\partial^2 T}{\partial r^2} + \frac{2}{r}\frac{\partial T}{\partial r}\right) + \begin{cases} Q e^{-\lambda t} & (r_1 \le r \le a) \\ 0 & (0 \le r \le r_1) \end{cases} \qquad (13)$$

with boundary conditions

$$T(a,t)=0,\ T(r,0)=0,\ T(0,t)<\infty \qquad (14)$$

The solution for (13), (14) has been found by G. Comenetz, but he never published his derivation. I have reworked the problem, and the solution, which is too long to give here, will be published in the chapter on the Earth's interior by Dr. J. A. Jacobs to appear in volume 47 of Springer's *Handbuck der Physik*. The result obtained is

$$\frac{T}{Q/\rho c}=\frac{ae^{-\lambda t}}{\lambda r\sin\eta a}\begin{cases}\sin\eta r\left(1-\dfrac{\sin\eta h}{\eta a}-\dfrac{r_1}{a}\cos\eta h\right) & 0\le r\le r_1\\ \sin\eta r-\dfrac{r}{a}\sin\eta a+\left(\dfrac{r_1}{a}\cos\eta r_1-\dfrac{\sin\eta r_1}{\eta a}\right)\sin\eta\,(a-r) & r_1\le r\le a\end{cases}$$

$$-\frac{2a^3}{\pi kr}\sum_{m=1}^{\infty}e^{-\frac{\pi^2m^2kt}{a^2}}\sin\frac{m\pi r}{a}\left[\frac{(-1)^m+\dfrac{1}{\pi m}\sin\dfrac{m\pi r}{a}-\dfrac{r_1}{a}\cos\dfrac{m\pi r_1}{a}}{m(a^2\eta^2-m^2\pi^2)}\right] \qquad (15)$$

where

$$r=\sqrt{\frac{\lambda}{k}}\ \text{and}\ h=a-r_1$$

The heat-flow is given by

$$F(r,t)=-\frac{K\partial T}{\partial r} \qquad (15A)$$

hence

$$\frac{F(r,t)}{H_o}=-\frac{ae^{-\lambda t}}{r\eta^2\sin\eta a}\begin{cases}\left(\eta\cos\eta r-\dfrac{\sin\eta r}{r}\right)\left(1-\dfrac{\sin\eta h}{\eta a}-\dfrac{r_1}{a}\cos\eta h\right) & 0\le r\le r_1\\ \eta\cos\eta r-\dfrac{\sin\eta r}{r}-\left(\dfrac{r_1}{a}\cos\eta r_1-\dfrac{\sin nr_1}{\eta a}\right)\left(\eta\cos\eta\,(a-r)+\dfrac{\sin\eta\,(a-r}{r}\right) & r_1\le r\le a\end{cases}$$

$$+\frac{2a^3}{\pi r}\sum_{m=1}^{\infty}e^{-\frac{\pi^2m^2kt}{a^2}}\left[\frac{(-1)^m+\dfrac{1}{\pi m}\sin\dfrac{m\pi r_1}{a}-\dfrac{r_1}{a}\cos\dfrac{m\pi r_1}{a}}{m(a^2\eta^2-m^2\pi^2)}\right]\left[\frac{m\pi}{a}\cos\frac{m\pi r}{a}-\frac{1}{r}\sin\frac{m\pi r}{a}\right] \qquad (16)$$

It should be noticed that both these equations have been taken in forms such that only the diffusivity, and not the conductivity, need be known for numerical calculations.

4. The Solution for the Non-Radioactive Earth

To complete the thermal picture of the Earth, the solution for the cooling of a non-radioactive Earth is required. The numerical results will depend, of course, upon the values adopted for the initial temperature distribution, which has been assumed to be everywhere the melting-point of the substance concerned—with a slight correction near the surface where the temperature is assumed to be zero. A smooth curve (see Figure 1) was drawn using the melting point values of Daly (1943) for the mantle and Simon's (1953) values for the melting point of iron at high pressures in the core.

The equation (6), with the conductivity K constant, reduces to

$$\frac{\partial T}{\partial t}=k\left(\frac{\partial^2T}{\partial r^2}+\frac{2}{r}\frac{\partial T}{\partial r}\right) \qquad (17)$$

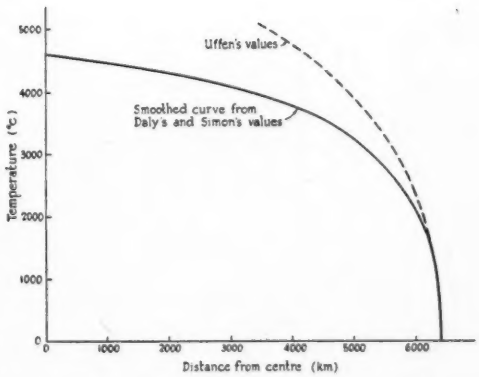


Fig. 1. Initial Temperature Distribution in the Earth.

with boundary conditions

$$T(a,t)=0, T(r,0)=f(r), T(0,t)<\infty \quad (18)$$

The solution, first given by Fourier, is

$$T(r,t)=\frac{1}{r}\sum_{m=1}^{\infty} A_m \sin\frac{m\pi r}{a} e^{-\frac{m^2 k \pi^2}{a^2}t} \quad (19)$$

where

$$A_m=\frac{2}{a}\int_0^a r f(r) \sin\frac{m\pi r}{a}\, dr \quad (20)$$

The heat flow, defined by equation (15A), follows at once. The A_m can be regarded as the coefficients in the expansion of the function $rf(r)$ in a Fourier sine series. Hence these coefficients can be found to any desired degree of accuracy by a harmonic analysis of the function $rf(r)$. For the present calculation, a sine series was evaluated to 44 terms. The results obtained for the temperature and heat flow at six depths and six times are listed in Table 1.

TABLE 1

Temperature and Heat Flow—Non-Radioactive Earth

Temperature (° C)

Time (10^8yr.)	Depth (km.) 50	100	500	1000	2000	2900
0	666	1220	2310	2856	3615	3958
·25	491	940	2296	2851	3613	3955
·5	409	796	2279	2848	3611	3953
1	328	647	2220	2839	3604	3949
2	256	510	2061	2818	3590	3942
3	220	439	1914	2786	3575	3934
4	197	394	1791	2742	3560	3926

Heat Flow (cals./cm.^{2}yr.)

Time (10^8yr.)	Depth (km.) 50	100	500	1000	2000	2900
·25	19·57	18·34	5·33	5·83	4·77	3·72
·5	16·50	16·35	5·69	5·85	4·87	3·82
1	13·36	13·81	7·15	5·86	4·95	3·93
2	10·50	11·15	9·36	6·12	5·04	4·15
3	9·03	9·68	10·27	6·71	5·10	4·33
4	8·09	8·71	10·55	7·44	5·16	4·48

5. Radioactivity in the Earth

The naturally radioactive substances of long period which occur in sufficiently high concentrations in rocks to have a decided thermal effect are U^{238}, U^{235}, Th^{232}, and K^{40}. The numerical evaluation of the solution given in section (3) requires a knowledge of the decay constant of each of these four radioactive substances, as well as their rates of heat production and their concentrations in the shells of whatever Earth-model is selected. In section (6), the Earth-models selected have shells of compositions similar to certain rock-types and meteorites. I have examined the literature of radioactive determinations with care in order to choose these values, having in general been guided by the principle that of two determinations the lower is the more likely to be right, since determinations of trace elements until recently were very inaccurate, and tended to err on the high side. For example, Ahrens, *et al.* (1952) recently found that older determinations of the concentrations of potassium in ultrabasic rocks and dunite were twenty to one hundred times too high.

6. Earth-models

Many different writers have given models for the Earth, in which they assign the composition of rock-types found on the Earth's surface to the spherical shells within, the properties of which are known only from seismic evidence. Two such models have been used here. The first is a simplified one, in which the crustal layers are replaced by a single layer 20 km. thick of granite-granodiorite composition, the mantle is identified as dunite, and the composition of the core is taken as that of iron meteorites. Apart from the simplified crustal layer, this model is identical with that proposed by Bullen (1947).

The second model, due to Adams and Williamson (1925) is more detailed. A structure of four 15 km. layers is taken for the crust, with the top layer of granite-granodiorite composition, and the remaining three layers of composition changing from acidic to basic. From 60 to 1600 km. the material of the mantle is identified as dunite, and from 1600 to 3000 km. it is taken as of the same composition as pallasitic meteorites. The core again is identified with iron meteorites. This model is especially valuable because the number of shells will make it possible to fit almost any other type of model which may be proposed later. (See Table 2 on next page).

TABLE 2

Summary of Data for the Models

Depth (km)	Shell	Density		Initial Heat Production (10^{-6} cals./cm.^{3}yr.)			H_0	$H_0/\rho c$ (10^{-6} degrees/yr.)		
		K		U^{238}	U^{235}	Th.	K	U^{238}	U^{235}	Th
Model I :										
0–20	granite-granodiorite	2·76	15	11	14	7·1	24	18	23	12
20–2900	dunite	4·77	·0099	·083	·11	·056	·0071	·059	·079	·040
2900–6371	iron (meteorite)	10·70	·0081	·046	·060	·031	·0074	·042	·055	·028
Model II :										
0–15	granite-granodiorite	2·8	15	11	14	7·1	24	18	23	11
15–30	changing from	2·9	6·9	5·4	7·1	3·1	8·8	6·9	9·1	4·0
30–45	acidic to	3·1	6·1	3·7	4·8	2·3	6·1	3·7	4·3	2·3
45–60	basic	3·3	3·0	2·6	3·2	1·6	3·1	2·7	3·3	1·7
60–1600	dunite	3·9	·0085	·069	·091	·046	·0057	·046	·061	·031
1600–3000	pallasite	7	·013	·059	·076	·039	·0087	·039	·051	·026
3000–6370	iron (meteorite)	10	·0072	·043	·055	·028	·0072	·043	·055	·028

The symbols Q_0 and H_0 have the same meaning as in section 3, except that the subscripts here mean that these are for time zero, *i.e.* 4×10^9 years ago.

7. Results and Conclusions

I have applied the solution of section (3) to the data of the following sections to calculate the temperatures and heat flows due to radioactive heating at six depths in the Earth and six times in the past. The computation involved was extremely laborious, even for this limited number of cases, and would have been well-nigh impossible without the aid of an electronic digital computer.

The actual temperatures and heat-flows in the two models are given in Table 3. The present surface heat flow for the two models can readily be calculated, and is 39 cals./cm.^{2}yr. for model I, 53 cals./cm.^{2}yr. for model II. The observational results in section (3) show that model II in all likelihood contains too much radioactivity in its upper layers. Fortunately the computations have been set up in such a manner that the effect of introducing different concentrations in the same shells can be ascertained relatively easily. I hope, therefore, to examine further models in the near future.

The results already obtained are, however, of great interest. They show, for example, that the Earth has probably been heating up at great depths throughout its history, and that near-surface thermal conditions were very different from what they are today 4 billion years ago. Any

TABLE 3

Total Temperatures and Heat Flows in the Models

Temperatures for model I : (° C.)

Time (10^9yr.)	Depth (km.) 50	100	500	1000	2000	2900
0	666	1220	2310	2856	3615	3958
·25	905	1216	2339	2894	3656	3992
·5	861	1166	2361	2930	3693	4024
1	763	1056	2382	2988	3753	4077
2	606	892	2358	3072	3844	4162
3	521	785	2308	3124	3913	4225
4	461	714	2248	3163	3970	4282

Temperatures for model II : (° C.)

Time (10^9yr.)	Depth (km.) 50	100	500	1000	2000	2900
0	666	1220	2310	2856	3615	3958
·25	1447	1520	2330	2885	3643	3985
·5	1424	1576	2346	2910	3667	4010
1	1267	1481	2369	2960	3705	4052
2	997	1245	2383	3017	3763	4117
3	825	1073	2347	3057	3808	4168
4	712	934	2298	3086	3841	4209

Heat flows for model I : (cals./cm.^{2}yr.)

Time (10^9yr.)	Depth (km.) 50	100	500	1000	2000	2900
·25	15·1	14·5	5·3	5·8	4·8	3·3
·5	15·2	14·5	5·7	5·8	4·9	3·2
1	14·2	13·8	7·0	5·9	5·0	3·2
2	13·1	13·0	9·1	6·1	5·0	3·3
3	12·3	12·5	10·4	6·7	5·1	3·5
4	11·6	11·8	11·1	7·4	5·1	3·6

Heat flows for model II : (cals./cm.^{2}yr.)

Time (10^9yr.)	Depth (km.) 50	100	500	1000	2000	2900
·25	12·0	5·4	5·3	5·8	4·8	3·7
·5	14·8	7·1	5·6	5·8	4·9	3·7
1	16·3	9·5	6·6	5·9	4·9	3·8
2	15·8	11·1	8·2	6·1	5·0	3·9
3	14·7	11·2	9·3	6·6	5·0	4·0
4	13·7	11·0	10·9	7·2	5·1	4·2

further interpretation of the results would be beyond the scope of this essay, but there does seem to me to be definite hope of settling the controversial question of whether or not the picture of an Earth cooling by conduction alone can satisfy geological evidence.

NOTE.—Because the space available for this essay was limited, some of the tables included in it had to be omitted, as well as the list of references. The complete essay may be inspected or borrowed from the offices of the British Association, and the author can supply reprints in complete form to anyone who writes to him at the Geophysical Laboratory of the Department of Physics, University of Toronto.

JOHN PHILLIPS AND THE EARLY MEETINGS OF THE BRITISH ASSOCIATION[1]

by

J. M. EDMONDS and P. A. BEARDMORE

Professor John Phillips, F.R.S. (1800–1874), was the nephew, ward and amanuensis of William Smith, 'the Father of English Stratigraphical Geology'. He earned fame as a palaeontologist and geologist, publishing volumes on Yorkshire, South-west England, the Malvern area, and, later in life, on the geology of the Oxford district and on Vesuvius, and although his formal education ended before he was fifteen years of age, he was successively Professor of Geology at King's College, London; at Trinity College, Dublin; and for twenty years at the University of Oxford. For seven years he served with the Geological Survey, and at Oxford he became the first Keeper of the University Museum.

Phillips was a secretary of the Management Committee at York for the inaugural meeting of the British Association in 1831. He was appointed assistant secretary at Oxford the following year and continued as assistant general secretary until 1862. He was President of the Association for the Birmingham meeting in 1865 and was a vice-president and an auditor at the time of his death in 1874. In addition to his executive duties, Phillips served on special sub-committees of the Association, contributed numerous scientific papers on geology, astronomy and meteorology to its meetings and was editor of the Annual Reports.

The scientific record and achievements of the British Association are contained in this succession of reports, and in the present paper emphasis will be placed rather on the incidental and informal happenings which do not readily find a place in the official records. Phillips's correspondence, including a series of letters written by him to his sister, Ann Phillips, is preserved in the Department of Geology, Oxford University, and the present head of the department, Professor L. R. Wager, Sc.D., F.R.S., has kindly allowed this material to be consulted and used for the present purpose.

John Phillips began his career when he was appointed Keeper of the Yorkshire Philosophical Society's Museum at York towards the end of 1825. Previously, he had travelled the country widely in the company of his uncle, William Smith, helping him in his work as a 'mineral surveyor' and 'civil engineer' and assisting him in the compilation of data on the stratigraphy of England, which was published as a series of county geological maps. These travels included a number of visits to Yorkshire, and in 1824 both uncle and nephew began permanent residence in that county, following an invitation extended by the Philosophical Society to Smith to give a series of lectures in York.

As secretary of the Yorkshire Philosophical Society Phillips received a letter, dated February 23, 1831, from David Brewster, editor of the *Edinburgh Journal of Science* (quoted in Howarth, O. J. R., 1922, *The British Association—A Retrospect*, p. 13), proposing the establishment of a 'British Association of Men of Science', and asking whether York, 'as the most central city in the three kingdoms', would be able and willing to accommodate the first assembly. This proposal was the result of discussions among a group of scientists who had expressed publicly their anxiety concerning the state of scientific endeavour and achievements in this country; the decline in the status of the

[1] Shortened version of a paper read before the Assembly of Corresponding Societies at the Oxford meeting of the British Association on September 8, 1954.

further interpretation of the results would be beyond the scope of this essay, but there does seem to me to be definite hope of settling the controversial question of whether or not the picture of an Earth cooling by conduction alone can satisfy geological evidence.

NOTE.—Because the space available for this essay was limited, some of the tables included in it had to be omitted, as well as the list of references. The complete essay may be inspected or borrowed from the offices of the British Association, and the author can supply reprints in complete form to anyone who writes to him at the Geophysical Laboratory of the Department of Physics, University of Toronto.

JOHN PHILLIPS AND THE EARLY MEETINGS OF THE BRITISH ASSOCIATION[1]

by

J. M. EDMONDS and P. A. BEARDMORE

PROFESSOR JOHN PHILLIPS, F.R.S. (1800–1874), was the nephew, ward and amanuensis of William Smith, 'the Father of English Stratigraphical Geology'. He earned fame as a palaeontologist and geologist, publishing volumes on Yorkshire, South-west England, the Malvern area, and, later in life, on the geology of the Oxford district and on Vesuvius, and although his formal education ended before he was fifteen years of age, he was successively Professor of Geology at King's College, London; at Trinity College, Dublin; and for twenty years at the University of Oxford. For seven years he served with the Geological Survey, and at Oxford he became the first Keeper of the University Museum.

Phillips was a secretary of the Management Committee at York for the inaugural meeting of the British Association in 1831. He was appointed assistant secretary at Oxford the following year and continued as assistant general secretary until 1862. He was President of the Association for the Birmingham meeting in 1865 and was a vice-president and an auditor at the time of his death in 1874. In addition to his executive duties, Phillips served on special sub-committees of the Association, contributed numerous scientific papers on geology, astronomy and meteorology to its meetings and was editor of the Annual Reports.

The scientific record and achievements of the British Association are contained in this succession of reports, and in the present paper emphasis will be placed rather on the incidental and informal happenings which do not readily find a place in the official records. Phillips's correspondence, including a series of letters written by him to his sister, Ann Phillips, is preserved in the Department of Geology, Oxford University, and the present head of the department, Professor L. R. Wager, Sc.D., F.R.S., has kindly allowed this material to be consulted and used for the present purpose.

John Phillips began his career when he was appointed Keeper of the Yorkshire Philosophical Society's Museum at York towards the end of 1825. Previously, he had travelled the country widely in the company of his uncle, William Smith, helping him in his work as a 'mineral surveyor' and 'civil engineer' and assisting him in the compilation of data on the stratigraphy of England, which was published as a series of county geological maps. These travels included a number of visits to Yorkshire, and in 1824 both uncle and nephew began permanent residence in that county, following an invitation extended by the Philosophical Society to Smith to give a series of lectures in York.

As secretary of the Yorkshire Philosophical Society Phillips received a letter, dated February 23, 1831, from David Brewster, editor of the *Edinburgh Journal of Science* (quoted in Howarth, O. J. R., 1922, *The British Association—A Retrospect*, p. 13), proposing the establishment of a 'British Association of Men of Science', and asking whether York, 'as the most central city in the three kingdoms', would be able and willing to accommodate the first assembly. This proposal was the result of discussions among a group of scientists who had expressed publicly their anxiety concerning the state of scientific endeavour and achievements in this country; the decline in the status of the

[1] Shortened version of a paper read before the Assembly of Corresponding Societies at the Oxford meeting of the British Association on September 8, 1954.

scientific worker; the attitude of indifference of the Government to scientific effort and to scientists; the failure of the major scientific societies to encourage research; and finally the unsatisfactory position regarding patent law. Brewster himself was foremost in outspoken criticism on these points. He asserted that only by the action of scientists themselves would any change be brought about in the attitude of the public towards these problems. A group of scientists, centred on Edinburgh and including Brewster, became interested in a development in Germany where a Congress of Scientists had been held annually since 1822. This meeting was held in a different city on each occasion and, beginning in modest fashion, it had quickly attained a national reputation as a forum for scientific opinion. An account of the meeting held in Hamburg in 1830, written by J. F. W. Johnston, was published in the *Edinburgh Journal of Science*, 1831, pp. 189–244, and it was believed by the sponsors of the British proposal that this German assembly would serve as an appropriate model for a gathering of scientific men in Britain.

Dr. Brewster's letter was brought before the Council of the Yorkshire Philosophical Society[1] at its meeting on February 28, 1831, and it was resolved

> that the Secretaries be instructed to communicate to Dr. Brewster that the Council will be happy to concur in his suggestion and to contribute all in their power to promote the objects of the scientific association as proposed in his letter.

Although John Phillips did not have any part in originating the idea of a 'British Association of Men of Science', he at once became the principal agent in the movement to bring the meeting about. Acting under the instruction set out in the minute above, he was in frequent communication with the Edinburgh founders during the spring of 1831. Letters written to Phillips by John Robison, secretary of the Royal Society of Edinburgh, indicate that there were early doubts on whether the proposal should be followed up:

> I find a very general disposition to go into the plan, but considerable doubts as to whether it should have been begun this year in the present excited state of feeling on other subjects[2].
>
> (Letter dated March 25, 1831.)

On April 29, Robison asked Phillips to find out from the Council of the Yorkshire Philosophical Society whether they thought that the meeting should be deferred until the summer of 1832. This letter was delivered to Phillips in London, where he was giving a series of lectures on geology at London University (University College). He was away from York for about two months but his presence in London enabled him to arouse interest amongst scientists living in the South of England, and he showed no willingness to acquiesce in Robison's suggestion of a postponement.

On May 26, Robison again wrote:

> I heard a few days ago incidentally that you had been for some time in London, & learned at the same time that several friends of Science there had spoken of their intention of attending the meeting at York. I am happy to have this confirmed by you and trust that after all the dubity which has been felt, a fair attendance may still be looked for. Much will however depend on the efforts of such men as Mr. Murchison, and I should have better expectation of the association finding general favour if the promotion of the plan appeared to come from London rather than from Edinburgh....

Phillips was quick to act on this suggestion and a circular was issued from London at the beginning of June announcing the meeting. This prompted Robison to write on June 8:

> I have had the pleasure of receiving your letter with the packet of circulars by which I am glad to see I am relieved from any further interference in the affair of the Meeting. It might have been as well not to have addressed the circular from London any more than from Edinburgh, as York should be looked on as head quarters until the persons who may assemble shall frame their own regulations....

Both John Robison and J. F. W. Johnston continued to contribute suggestions to Phillips during the following

[1] We are greatly indebted to Mr. S. Melmore for making extracts from the Minute Book of the Society and also to Mr. G. F. Willmot, Keeper of the Yorkshire Museum, who gave permission for this to be done. Mr. Melmore has also kindly provided the extracts from the *Yorkshire Gazette*.

[2] i.e. The Reform Bill agitation.

weeks, regarding details of organisation of the meeting, based on Johnston's experience of the assembly of the German society at Hamburg.

Soon after Phillips's return from London, the council of the Yorkshire Philosophical Society set up a committee[1]

> to make the necessary arrangements for the reception of the Strangers coming to York at the ensuing Scientific meeting with power to add to their number, & that they be authorized to announce the meeting in the name of the Council.

(Minute of the Council of July 4, 1831.)

On July 12, the committee issued a lithographed circular letter announcing the meeting in the name of the Society, signed by the chairman of the committee, the Rev. Wm. Vernon Harcourt, and the secretary, John Phillips, and on September 7 a further announcement was circulated. A minute of the council of the Society for September 22 reads :

> The Chairman having produced a plan for the constitution & conduct of the proposed British Association for the Advancement of Science, Resolved that the Plan be adopted by the Council, and that the Chairman be requested to propose it for the consideration of the General Scientific Meeting.

Although there had been a reasonably encouraging response to the circulars announcing the meeting, there was considerable doubt up to the last moment as to the success of the gathering because of the antagonism which Brewster had aroused among numerous professors and other recipients of scientific honours by making a generalised attack on their competence. A letter to Phillips, from J. D. Forbes, of Edinburgh, dated August 3, 1831, refers to this matter :

> I had a letter the other day upon various topics from Professor Whewell of Cambridge and in mentioning Dr. Brewster's attack upon Professors in their scientific character (which his best friends must allow to be unadvised & most unreasonable) he adds that even if other circumstances permitted he should certainly not think of going to York, *to rally under Dr. Brewster's banners*. Now this is truly too absurd.

[1] The Minute Book of this committee is now in the possession of the British Association.

Phillips's own anxiety is recorded in the following description :

> on the Sunday afternoon before the Monday on which the Meeting was to commence, he walked after Service in the Museum Garden, for the purpose of meeting and welcoming any strangers, and found for some time no one had arrived. At last he discovered four persons, all friends of his own, who had come from a considerable distance to lay the foundation of the Association. These four persons, eminent as they were, would not be sufficient to make a meeting ; and Professor Phillips had anticipated seeing many more illustrious strangers when he commenced his walk.

(Quoted from Fison, Mrs., 1859, *Handbook of the British Association.*)

On Monday, September 26, on the eve of the meeting, there was a conversazione at the Museum.

> The rooms were lit by gas whose brilliant light reflected every object they contained ; and as they were paraded by groups of elegant females, the charms of beauty and the varied stores of philosophy seemed united to attract the most careless observer,—and could not fail to interest and please. . . . At ten Mr. Phillips delivered an animated lecture on the Geology of Yorkshire—got up on the spur of the moment ; without any premeditation, and which showed the complete mastery he has of the science. . . . The company were delighted with the lecture and separated about eleven o'clock.

(Extract from the *Yorkshire Gazette* for October 1, 1831.)

On the following day, Tuesday, September 27, Lord Milton, the President of the Yorkshire Philosophical Society, was installed as chairman, and the Rev. Wm. Vernon Harcourt said :

> This meeting owed its origin to some distinguished cultivators of science many of whom were now present, who thought that great benefits would result to science in Great Britain, by establishing meetings similar to those on the continent ; and he could not satisfy himself, without naming the gentlemen who were the first projectors of the scheme. Dr. Brewster, Mr. Robison, Mr. Johnston, & Mr. Murchison were among its first & most assiduous promotors. The Council of the Yorkshire Philosophical Society fully concurred. . . .

Phillips then described the steps taken by the Yorkshire Philosophical Society—

and the wide distribution of the circular. Harcourt then spoke again :

It would be proposed to the meeting, to form a British Association for the Advancement of Science, having for its objects,—1st, to give a stronger impulse and more systematic direction to the progress of science ; and 2dly, to direct the national attention to scientific objects. In advocating this new association, he did not agree with those who thought that science had declined : he did not believe that it was under a declension : at no period, with the exception of Newton, were there names that ranked higher than those of Davy, Wollaston, and Young : and there were men still remaining among us, whom he must not name, but whom he was happy to see present at that meeting. . . .

(Extracts from the *Yorkshire Gazette* for October 1, 1831.)

The full official account of Vernon Harcourt's speech is contained in the published proceedings of the York meeting issued in February 1832 as the ' First Report of the British Association for the Advancement of Science '.

When the York meeting was over, John Phillips was left with three tasks : first of issuing the Report, then of encouraging the preparation of the next meeting, at Oxford, and finally of reconciling the divergent views of the principal promotors. On October 17, only a fortnight after the meeting ended, J. D. Forbes wrote to him from Edinburgh :

The hands into which the affairs of the Association were committed, left no doubt upon the minds of anyone that they would be managed in the manner most conducive to the interests of the body. I may be allowed however to suggest that it is of great consequence that no delay should exist in the transmission of the short statement & more extended report which you have in preparation. Any feeling which the York Meeting excited will rapidly die away, & it is of the utmost importance that the nature & objects of the permanent Association should be known before the recollection of their first act is effaced.

Phillips went on with the task without delay and the preliminary leaflet was soon issued. A letter dated November 12, from Professor Adam Sedgwick, of Cambridge, acknowledges this :

I have this moment rec^{d} the " Objects and Rules of the British Association for the promotion of Science ", and I enclose 5£ as my composition in lieu of Annual subscriptions. It would have delighted me above all expression to have been present at the last meeting but it was impossible without completely destroying my summer's work—Murchison ran away in June, but left me with a set of detestable MS.S. to pass thro' the press, which kept me in Cambridge to the end of July !

After referring to his field work in Wales for the rest of the summer and to his programme there for 1832, he adds :

Under these circumstances I almost fear that I shall not be able to find my way to the next meeting at Oxford.

Fortunately this fear was unfounded. Sedgwick attended the Oxford meeting and there accepted the office of President elect for the Cambridge meeting for 1833.

With regard to the ultimate objects of the Association there remained a divergence of opinions which only time would solve. David Brewster expressed his strong views in a letter to Phillips dated November 19, 1831:

I may mention to you *confidentially* that in writing yesterday to Mr. Harcourt I stated to him the risk to which the Association is exposed by the introduction into his address of the controversial parts respecting the non-decline of Science, & by the circumstance of both him and Lord Milton having reprobated the direct *encouragement of Science* by the government. These things I had entirely forgotton, tho' I felt their inconvenience at the time ; but when I was in Edin. last week, there was put into my hands a London Magazine taunting me by name for having heard Mr. Harcourts observations without replying to them, and on the same evening I received a Proof, that of Mr. Johnston's account of the York Meeting, in which without my communication with any person whatever he has felt it his duty to reply to Lord Milton's observations. These I believe will be found to be the reasons why Babbage & South will not join our Committees, & if something is not done to prevent it, our Association will find its grave in Oxford. Herschel, Babbage, South, Leslie, Henry, Harvey, Johnston, Brown, are unanimously of opinion that the Mathematical & Physical Sciences (not the Natural History Sciences) are rapidly declining, & can only be saved by the direct encouragement of the Government ; and if the Pres. & V. Pres. of the Brit. Assoc. avow opposite opinions, on topics on which they are unable to judge, & publish these under the sanction of the Association, all the above individuals, and the crowds who think with

them will withdraw from the Association. I never will withdraw on any such grounds, but if the officebearers of one year attack opinions supported by *nine tenths* of the Association, the office-bearers of another year will be compelled to defend their opinions, and the House thus divided against itself will not long stand. . . . I have no doubt that this has taken place from the good motive of having the cooperation of Whewell who is crazy about the decline of Science, and hostile to Government doing anything for it. But if *one* Whewell is gained *twenty* are lost to the Association, and one of its leading and grandest objects sacrificed for ever.

For its second meeting, in 1832, the Association was invited to Oxford. C. G. B. Daubeny took the responsibility of issuing the invitation and of making the arrangements. He wrote to Phillips on November 12, 1831 :

Buckland has announced the intended Meeting to the Ashmolean Society, and I think I may say from the disposition to promote the plan evinced both here and at Cambridge we are sure of a good attendance, if Cholera or Politics do not interfere. As the Meeting will probably take place in Term time, the Commemoration happening too late to allow of our postponing it till then, our accommodation within the Colleges will be limited, but I am sure that everything will be done in other respects to give our guests an hospitable reception. In all cases you may depend on a bed being reserved for yourself at the Museum in Broad Street. My Introduction to the Atomic Theory being published, I take the opportunity of sending a Copy for the Institution at York. . . .

Phillips's correspondence shows that from this time forward he was rarely free from the business of the Association. He was, however, evidently prevented by illness from attending the Oxford meeting, and this must have been the more disappointing because during the meeting his uncle, William Smith, was handed the first Wollaston Medal award of the Geological Society of London. Having been given a constitution at York, the Association developed into an organisation at Oxford, where, according to Murchison,

under the presidency of Buckland, the body was then licked into shape, and divided into six sections.

(Geikie, A., 1875, *Memoir of Sir Roderick Murchison*, Vol. 1, p. 204.)

At the same time, the Rev. W. V. Harcourt was chosen as General Secretary and John Phillips was appointed Assistant Secretary. At the beginning of July 1832, Phillips was in Manchester giving a course of zoological lectures, and writing to his sister, he said :

As the intervals [between lectures] are so long, I can be able to attend to the Oxonian Report.

At Oxford, and also in the following year (1833) at Cambridge, the arrangement of the meeting was largely in the hands of the local secretaries and it was simplified by the existence of ample lecture rooms and the presence of numerous persons interested and experienced in the objects of the Association. This was also the case in 1834 in Edinburgh, and during the spring of that year Phillips was corresponding with the local secretary, John Robison, who had contributed so largely to the successful organisation of the first meeting. Robison was in favour of a simplification of the arrangement of meetings, but there was one suggestion made by Phillips with which he found himself unable to agree, and on April 6, 1834, he wrote :

You have too favourable an opinion of Scottish zeal and industry when you propose to rouse a Mathematical Section to work at 7 A.M. We cannot depend on regular work commencing any day before 11. Committees and persons charged with specific duties, may meet at 10 but even that would require an effort, as many of the persons who will be efficient figures in their numbers, will have to give breakfasts, and attend to their guests. From half past ten or eleven A.M. to 5 P.M. will be a good position of time for the Sections to work, *if they be allowed to work all that time, without being distracted from their business, or called away to other meetings.* If on the contrary there be a continual changing of employment, by alternating Sectional and General Meetings on the same mornings, little business will be got thro', and uncertainty & confusion will ensue. Philosophers are not always good men of business, and lookers on sometimes see more of the game than those who play it, whence I would say that the remarks which were made by lookers on at Cambridge merit some attention. I feel strongly impressed by some of the observations in the public papers at the time of that meeting, that the general plan requires to be very much simplified in order to give that clear conception of it to everyone taking a part

in the proceedings to enable them to go thro' with it without needless efforts of memory or constant consultation of a programme. The history of the German Association confirms this view, and shows the advantage of not interrupting the morning labours of the Sections, where after all, the most important part of the work must be done.

The success which the Association had so rapidly achieved increased the need for careful planning and organisation. Towards the end of the Edinburgh meeting Phillips wrote to his sister :

I am yet in a state of hurry & confusion such as I never knew & this must account for my silence during the past tremendously exciting week. I have been hard worked and have done I hope well. At the last general committee meeting my salary was raised to £200 on certain conditions & I must now view myself as bound to the Association by a stronger and more responsible bond.

(Letter dated September 15, 1834.)

Following this successful meeting in Scotland, that for the next year, 1835, was arranged in Dublin. On July 23, Phillips began his journey there by travelling from Manchester to Liverpool by the railway which had been opened in 1830.

Entered the Train & the Railway & fly Jack & begone. In 80 minutes we reached Liverpool 32 miles, & soon by another omnibus were landed at our several hotels. My dear Annie [he wrote to his sister], you must certainly come to feel the *strange impression* of this flying Steed of Iron. It does so hurry and so flurry on, you shake & sleep & start & wonder at the gliding Houses trees & Churches,—the trains which meet & pass you like the swiftest birds with a rushing sound of the master power (Steam) & a confused picture of colours & forms not at all distinct as men women carriages &c that it is all like magic & can not be understood by a mere description. Then you are dragged through a tunnel full of gas lamps, then laid hold of by ruffian porters & crammed into an omnibus whether you will or no & whirled away the man who guides (only) knows whither.

While in Liverpool, Phillips took the opportunity of inspecting the accommodation for the meeting to be held there in two years' time (1837), finding the rooms ' abundantly sufficient '.

In Dublin, Phillips stayed in Trinity College with Professor Humphrey Lloyd, Professor of Natural and Experimental Philosophy. It was the friendship that developed between these two men which undoubtedly led Phillips to return to Trinity College as Professor of Geology nearly ten years later. Lloyd was the local secretary for this meeting of the Association,

and as he is an excellent man of business my labours are not very heavy. Yet, as they are of a kind to make my presence in Dublin necessary all the preliminary days, I have only had one trip to geologize in the vicinity.—the generosity of the Irish Members is extreme—nearly £1000 has been raised to make the cost of their visit as light as possible to the strangers, who will have only 5/– to pay at the ordinary (wine included). . . . The admissions are regulated according to a plan of mine which seems to leave little room for confusion or delay. The General Committee has red, the foreigners have blue, the Irishmen yellow and others white Tickets. On the Platform (that source of jealousy) we shall admit only Officers and Councillors—Ladies will be admitted by Ticket. Tickets for Ladies given only to strangers & to Irish members who subscribe to the Local Fund. There will be promenades, a Military Band, open gardens and gay refreshments. Three or four public dinners & breakfasts, a trip on the Railway &c.

(Letter to Ann Phillips, August 1, 1835.)

In a later letter he gives further details and lists the ' Stars ' (as he calls them) :

The attendance from Edinburgh is not very large. Jameson, Forbes, Robison, Abercrombie, Brewster all absent—Airy, Peacock & Challis of Cambridge all absent—Buckland absent from Oxford. Brown, Lindley, Faraday have not come from London,—Hooker has not come from Glasgow. We have Greenough, Sedgwick, Murchison, Taylor—Rigaud, Henslow, Whewell,—Powell, Greswell, Daubeny,—Alison, Traill, Johnston,—Sir. T. Brisbane, Sir J. Franklin, W. Snow Harris, Fox of Falmouth, Woolcombe of Plymouth &c.

When the meeting was over, Phillips wrote again to his sister :

Everything has worked very well, each Section thinks itself to have been eminently successful. . . . The weekly festivities were incredibly beautiful particularly the Zoological Fete, & (as I hear for I could not go to it) the Dublin Garden Fete. The most elegant dinner I ever saw was given to 300 members on Saturday in Trinity College, when the fine quire of the College sang delightfully. . . .

He went on to recount his visit to Florence Court as guest of Lord Cole :

Sedgwick and Murchison made me stop with them en route to see a limestone quarry

& a funny trip it was. The car (a jaunting car which by the by is a capital thing for geologizing) was drawn by a tired horse, which in this warm weather would stop at the whiskey houses for water. In one case the driver contested this point and the beast would not stir a step forward, so he turned her round and backed her for 100 yards & still he could not conquer but was obliged to give the water. Sedgwick had cigars, & gave one to the driver who after a while threw it down amongst the luggage & set fire to the things. At last driven to search by the strange smells we found Sedgwick's fossil gathering bag burnt & his umbrella reduced to a tatterdemalian thing, fit only for a ragged Irishman to bear. So he gave it to the unhappy driver who with all possible *insouciance*, laughed at the whole adventure.

In 1836 the Association visited Bristol and John Phillips arrived there to help in the final arrangements early in August. He stayed at the house of William Sanders and they had several geological excursions together, which included Banwell Cave and Cheddar Cliffs. Phillips was quite confident about the arrangements for the meeting and wrote :

Today the local Council meets to sanction programme & adopt final rules as to Ladies Tickets etc. I have no difficulty in the arrangements & can promise a most splendid meeting (as to Science). Sir D. Brewster—Sir W. Hamilton—Lubbock, half Cambridge—half Oxford—will be there. There is I think a probability that it *will be the best meeting yet held.* The Ordinary, Section rooms, &c. &c. are all provided—I have arranged a far easier plan of passing Members & all promises to go on very well. Thus you see I ought to be well satisfied & so I am. . . . You must attend the Meeting, for it ought to be superb, but I fear you may hardly get to the Sections. If any one else does you shall also. . . .

(From a letter to Ann Phillips, August 11, 1836.)

Evidently the Sections were not opened to women and among those excluded was a daughter of Bristol, Mary Carpenter, the pioneer in providing schools and homes for destitute and neglected children. At the British Association's meeting in 1860 at Oxford she was to read a paper before a crowded gathering in the Statistical Section on the need for Government help in dealing with the problem of those unfortunate children, but in 1836 she had to be satisfied with meeting the visiting scientists at evening meetings and receptions and at the private occasions of hospitality in her father's house.

The efforts of Phillips to make the registration of members as straightforward as possible were appreciated by at least one newcomer to the meetings of the Association as is shown by the following extract from the manuscript diary of John Eddowes Bowman, banker of Welshpool and Wrexham:[1]

August 21st. (Bristol). . . . I then went with Eliza to Professor Phillips [at Wm. Sanders' house] ; he was out but we saw Miss Phillips. I then went to procure my admission ticket, which I gained at once, though a perfect stranger and not a member previously, by saying I was a member of the Linnean Society and the author of papers published in its Transactions. I also gained much local information.

August 22nd. Today (Sunday) I have seen Phillips who wishes my paper on The Bone Caves to come on tomorrow, as he says there is always less to be done on the first day. . . .

This diary gives an excellent word-picture of the Bristol meeting and projects a vivid description of the sense of bustle and urgency, purpose and expectancy which are characteristic of the best meetings of the British Association throughout its history. Of a popular meeting in the Geological Section, he says :

. . . I hastened thither, and finding I could not get within three or four yards of the entrance, made my way down to the private committee entrance, and at length succeeded in getting in amongst the great men and the speakers, where I had a very comfortable seat, and heard and saw all. There was afterwards an excellent paper on the formation of mineral veins, and the change of metals in them by electricity. After dinner . . . I went to see the experiments of a Dr. Hare of Philadelphia on Voltaic Electricity ; and afterwards heard my friend Phillips, give an account of his views as to the blocks of granite from Shap Fells which are so copiously scattered over the eastern and southern counties for 80 or 100 miles. Afterwards Sedgwick, Murchison, Conybeare and De La Beche delivered their views, and Dr. Buckland summed up. This was a high treat, besides seeing many other distinguished men, among whom was Dr. Smith the first originator

[1] Copied and kindly placed at the authors' disposal by Mrs. H. L. Bowman and quoted by permission of the Rev. Sir Paget M. Bowman, Bart.

of Geological Maps ; a plain farmer-looking man.

With the successful completion of the Bristol meeting of 1836, it may be said that the pattern and tradition of the British Association was set. There might be anxieties through epidemics, civil unrest, venturings to remote or unsympathetic places, or opposition from antipathetic groups and organs of publicity, or curtailments because of wars and rumours of wars. Nevertheless, the form of the Association meetings was broadly fixed after this the sixth assembly. That it so quickly became a part of the scientific life of the United Kingdom, must in itself testify to the clear-sighted purpose of those early enthusiasts whose wishes were rapidly translated into effect by John Phillips. The Report of the Council presented to the General Committee of the British Association for the Advancement of Science in 1874 speaks of him in the following terms and serves as a fitting epitaph :

> In Professor Phillips, eminence in his own branch of Science and wide general culture were united with unselfish sympathetic nature, a genial kindly manner and with a singularly happy tact in the conduct of affairs. It was this rare combination of qualities which guided the Association through its early difficulties to the success it has at present achieved, and which now makes his loss felt as one which can never be filled up.

THE BIOLOGY OF SEWERS AND SEWAGE TREATMENT

At the Oxford Meeting of the Association, Section D (Zoology) heard five papers during the afternoon session of Tuesday, September 7, 1954, dealing with the Zoology of sewers and sewage treatment.

Professor E. A. Spaul, of the University of Leeds, opened with a description of the fauna of the sewage works, indicating its importance in relation to the efficiency and economy of treatment, and showed how variations occur with the type and flow of sewage, the medium in the beds, and seasonal change.

Mr. Tomlinson and Mr. Hawkes, of the Birmingham Corporation Sewage Works, followed with an account of the factors concerned with the size and balance of the insect population of percolating sewage filters, and methods of control.

Next, Dr. P. H. Silverman, of the Liverpool School of Tropical Medicine, discussed the dissemination of the eggs of the human tapeworm through their survival of sewage processing, and consequent distribution to pasturage where they infect cattle.

Then Dr. E. W. Bentley, of the Ministry of Agriculture and Fisheries, Infestation Control Division, dealt with the rat population in sewers, and the application of control measures based on knowledge of their behaviour.

Finally, Mr. W. B. Alexander, of the Edward Grey Institute of Field Ornithology at Oxford, surveyed the various types of birds attracted to sewage farms, and showed how they altered with the method of treatment at the works.

Professor Spaul has edited the papers, and they are printed in the order in which they were delivered.

Biological Problems of Sewage Treatment

by

Prof. E. A. Spaul

The efficient treatment and disposal of sewage is of primary importance to the health and well-being of modern communities with their complex organisation and varied activities. The application of scientific methods to its purification has become therefore an established practice of the sanitary authorities responsible for public health. Under natural conditions organic waste is reduced to simpler organic substances by bacterial action and eventually transformed into salts—such as nitrates, phosphates and sulphates needed by plants—by the activity of other bacteria in the presence of oxygen. These processes form the basis of any scheme for sewage disposal. The cesspits of the past and more recently the sewage farms where waste is discharged over land to be absorbed by the soil and rendered innocuous, are quite inadequate and unsuited to present-day needs created by industrial development and the spread of populated areas. In modern sewage treatment works there is a concentration and intensification of the natural processes so that effective purification by biological activity in an abundance of oxygen is produced in a few hours compared with days or months under ordinary conditions. The waste is conveyed by water together with surface drainage along the sewers to the disposal plant and after the removal of the heavy materials and solids, by screening and gravitation, the suspended organic matter is separated off in the sedimentation tanks. Here the regulation of the flow, aided when necessary by the addition of lime, brings about the settlement which forms the primary sedimentation sludge. The resultant sewage liquor passes on for biological oxidation in either the percolating filter where it is distributed by sprays from moving arms over beds containing a medium of stones or clinkers, providing plenty of air spaces and harbouring the active fauna, or through channels for the activated sludge process, when it is dosed with activated sludge and aerated by agitation with paddles or compressed air forced into the liquid. A final separation of flocculent material—humus or secondary sludge—occurs in each case during the passage of the oxygenated effluent through humus tanks, and at this stage a standard of purity is reached allowing the discharge of the effluent into a stream or river without causing pollution. The conditions in the sewage

system favour not only bacteria but a host of other organisms and with the plentiful and continuous supplies of organic matter as food and oxygen, moisture, adequate shelter throughout the year and limited fluctuations, dense populations can be supported whilst habitats suited to the association of a wide range of organisms are provided. Many species are present but only a limited number especially well suited to the somewhat unique and specialised environment have successfully established themselves in great numbers. The presence of this fauna is of great importance on account of its great influence both direct and indirect upon the working of the plant and the purification processes.

In the anaerobic conditions of the sedimentation tanks, fungi and protozoa exist with the bacteria, but in the percolating filters worms and insects in addition abound with some nematodes, molluscs, crustaceans and other arthropods, all thriving in the aerobic conditions and holozoic nutrition. There is not the same variety of species and numbers present in the humus tanks as the breakdown of the organic material is nearing completion. The most successful and important are species of earthworm, small enchytraeid worms, various species of crane, moth and hover flies, midges and protozoa.

It would appear that many of the species of worms and insects which have so effectively colonised the filter beds are derived from the fauna of mud flats in ponds and streams where there is a natural environment with features bearing a close resemblance to those of the beds. Access to the beds for these organisms is easy either in the sewage, through drainage or in the case of insects by flight.

The surface of the filter is covered with filamentous algae generally of the green or blue-green type which flourish in contaminated conditions and a slimy gelatinous zoogloeal growth of bacteria and fungi containing many other micro-organisms, mainly protozoa, covers the medium. Until this growth is established and matured the beds cannot function and purification can only occur while it is maintained. The sewage stimulates growth so that the algae spread and the film thickens and would fill the spaces in the medium and choke the bed, preventing the passage of sewage unless checked by the grazing activities of the worms and insect larvae living in the filter beds. Their constant feeding counters excessive growth, loosens and fragments the film causing older portions to be shed and fresh film to develop. In this way the beds are kept open and functioning and as they do not become clogged ' ponding ' is avoided. These activities are related to temperature so that there is a characteristic cycle in the growth. During the winter the film thickens and the surface growth becomes leathery but the greatly increased activity of the grazers in the spring produces a seasonal sloughing—the spring off-loading. There may be another in the autumn. Enchytraeid worms generally work at the surface but cold drives them down and it is their return that initiates the off-loading. Many follow the detached growth into the bed to break it up and feed on it. The large worms work near the bottom and many pass through with the effluent to drown and putrefy in the humus tanks. When this is excessive they become a nuisance and to relieve this situation the possibilities of their transfer to land for use in relation to problems of the soil might be worth exploration. In some works they come on to the surface especially on warm rainy nights, and may even leave the filter. This nocturnal activity does much to keep the filters open but if the migration is heavy difficulties arise. The distribution and density of the worms in the beds is therefore constantly changing and their movements are important in relation to the effective working of the beds.

The output of flies is enormous as conditions particularly favour more or less continuous and prolific breeding and provide, as in the case of worms, shelter from predators, but there is some degree of mutual destruction as many eggs and pupae and even larvae are eaten during grazing. Several generations of hundreds of eggs according to the species are produced during the year with scarcely a check in the winter, but the vagaries of the weather may affect production and survival, for the breeding times of the different species vary and such fluctuations will modify competition for food between them and produce variation in their relative numbers. The emerging flies are harmless but may reach such proportions at times that some appropriate treatment is needed to avoid a nuisance. The control must be achieved with the minimal disturbance of the beneficial activity of their larvae and the adequacy and balance of the fauna and hence the effective working of the filters. Much is known about the dosage and application required of some insecticides, and it would appear that the worms and protozoa, although susceptible, can tolerate the concentrations used. More information, however, is needed. Further, with this treatment the flora and fauna of the stream receiving the effluent have to be considered. Other methods, including biological control, are possible.

Protozoa play a particularly important part in the purification processes, especially in the activated sludge process. Their distribution through the system depends upon the type of food and amount of oxygen available, the

nature of the sewage and its distribution, and the macrofauna. In the sedimentation tanks there are anaerobic and saprophytic or holozoic forms with a wide range of physio-chemical tolerance. The filter beds are aerobic and nutrition ranges from saprophytic, through holozoic or parasitic, to holophytic with weaker sewage and seasonal variations influenced especially by the sloughing. In the humus tanks the protozoan fauna is scantier and shows seasonal variation similar to the beds. In the activated sludge channels the fauna is uniform throughout their length and dominated by the ciliates especially the stalked forms. The feeding habits of these forms with the secretion of mucus cause the flocculation of bacteria and the aggregation of particles which readily sink so that purification and clarification of the water are speeded up considerably.

It is evident that the individual and collective activities of the various groups of organisms have a definite relation to the efficient and economic working of the sewage system and their integration maintains a biological balance upon which depends an effective standard of purification. Although the fauna varies with the sewage works according to their construction, method of filtration, and location, and is affected by the nature and size of the medium of the bacteria beds, so that different organisms are dominant, the function is the same. The activity of the organisms in the fauna and their co-ordination, however, are influenced by the character and quality of the sewage (acidity, proportions of domestic and trade waste), its strength and the rate of flow so that adjustments and regulation are constantly needed to preserve an appropriate balance and obtain the best results. Sudden changes in quality or composition or the introduction of new chemicals or other substances into the waste can have a profound and prolonged effect upon the system and if as a result the beds become clogged, the flow must be stopped to allow the recovery of the fauna. In practice such stoppages can be expensive.

Ecological study supplemented by laboratory experiments have produced a wealth of information upon the functioning of the sewage processes but more is needed and investigations on the lines of Lloyd on insects, Reynoldson on enchytraeid worms, Barker on protozoa, and others in the West Riding of Yorkshire, extended to other areas would add substantially to our knowledge. The biological approach can contribute much of value to current problems concerned with improved performance of bacteria beds, the best combination of organisms for the most efficient treatment, population densities, and control. In fact, the realisation of the importance of the fauna to the successful operation of sewage and disposal plants has opened up a new field in applied biology and the study of the biology of sewage has therefore not merely scientific interest but great practical value and importance, for it has as its concern not only the effective management of sewage works but also its relation to the supply, use and control of water, water pollution, town planning and development, and agriculture.

The Insect Populations of Percolating Sewage Filters

by

T. G. Tomlinson and H. A. Hawkes

The percolating filter supports an unusually concentrated population of organisms covering a wide range of species. The function of the filtering medium is to provide a large surface for the growth of organisms responsible for the purification process. Hence numbers are related to the total area of the medium rather than to the superficial area of the filter. As an illustration of the relation, the total surface area of a cubic foot of filtering medium of 2-inch nominal size is about 20 sq. ft.

Sewage is usually applied to the filter at the maximum rate of flow compatible with the removal of some 90 to 95 per cent. of its organic content and also with the acitivity of the filter metazoa to keep pace with the growth of fungi and bacteria and the continual deposition of sewage solids. Since the efficiency of the filter decreases if the natural ventilation and free flow of sewage are impeded by excessive amount of organic matter, a favourable balance is essential.

Worms, springtails, and larvae of a number of species of flies feed on the organic film, breaking it down into faecal particles which are readily washed through the filter. The discharge of this material usually shows a marked seasonal rhythm as a result of the accumulation of organic matter during winter and its disintegration by larvae, during spring and summer.

The insect population belongs almost exclusively to the Collembola (springtails) and Diptera (flies), although the beetle *Cercyon ustulatus* is usually present in small numbers. A number of species of Collembola have been found but the most abundant is *Achorutes subviaticus*, the gregarious habit of which often produces a purple carpet on the surface of the filter. Compared with the filter flies, the development of *Achorutes* is slow, the average life during the summer being about 80 days.

The population of Diptera is dominated by one or more of some 8 to 10 species. These are :

Psychoda alternata, *P. severini*, *P. cinerea*, *Hydrobaenus* (*Spaniotoma*) *minimus*, *H. perennis*

Metriocnemus hygropetricus (*longitarsus*), *M. hirticollis*, *Anisopus fenestralis*, *Paracollinella* (*Leptocera*) *fontinalis*, *Scatella silacea*, *Spaziphora hydromyzina*.

The most prolific of the Diptera are *Psychoda alternata* and *P. severini*. During summer *P. alternata* completes its life cycle in about three weeks, a single female laying about 100 eggs. One individual *P. alternata* larva requires, on average, about 0·003 g. of solid matter in the form of film for complete development at 20° C. ; hence an amount of film equivalent to 10 lb. dry matter would provide sufficient food for the development of about 1½ million *Psychoda*. *Psychoda* breeds at all levels in the filter, its density being roughly proportional to the amount of film present.[1] It can breed successfully within the confines of the filter without emerging from the surface. At most works sewage is treated in circular filters dosed at short intervals from rotating distributors. Under these conditions film is concentrated in the surface layers of filtering medium and thus *Psychoda* flies are able to emerge from the surface in large numbers and may cause a nuisance. Various modifications in the operation of filters have reduced the total ouptut of flies by reducing the amount of organic matter, particularly in the surface layers of medium.

The 'window gnat' *Anisopus fenestralis* has similar habits to the psychodas, but the mature larvae usually migrate into a cleaner part of the filter to pupate. Unlike the pupae of *Psychoda*, which occur in the food material and can survive constant disturbance from feeding larvae, those of *Anisopus* are very vulnerable to suffocation. Hence in circular filters, where the film is uniformly distributed in the surface layers of medium, *Anisopus* cannot compete with *Psychoda* except when low spring temperatures give them a temporary advantage. *Anisopus* is particularly successful in rectangular filters where the widely-spaced distributor jets and relatively low concentration of film provide ideal conditions for pupation.[2]

The larvae of the chironomids (*Hydrobaenus* and *Metriocnemus*) are essentially grazers and are most successful when the filtering medium is covered only by a tenuous film. Under these conditions the larvae of *Psychoda*, which are adapted to feed immersed in their food, are at a disadvantage. Moreover, when food is scarce, the carnivorous larvae of *Metriocnemus* feed on the eggs and young larvae of *Psychoda*. *Hydrobaenus minimus* and *Metriocnemus hirticollis* mate only in a swarm when the temperature is above 10° C. Chironomid swarms are a common sight at some works but usually the flies do not disperse from the filters.

[1] Tomlinson, T. G., and Stride, G. O. *Proc. Inst. Sewage Purification*, 1945 (2), 140.
[2] Hawkes, H. A. *Ann. Appl. Biol.* **39**, 181, 1952.

Nuisance from *Psychoda* and *Anisopus* can be reduced by adding DDT or benzene hexachloride as a water-dispersible powder to the influent sewage.[3] Fortunately worms and springtails are not affected ; their numbers may even increase after treatment with insecticide in the absence of competition from insect larvae.

Dispersal of *Anisopus* has been studied by trapping flies in different directions at some distance from the filters. They tend to disperse in the direction of the wind but flight is inhibited by strong wind. The numbers increase with a rise in temperature and few flies leave the filters when the air temperature falls below 11° C. Dispersal is most active during spring and early summer, but the reasons for this seasonal variation in dispersal are not yet known.

The Survival of the Egg of the 'Beef Tapeworm,' *Taenia saginata*

by

Dr. P. H. Silverman

Man can acquire different tapeworms from various sources such as certain freshwater fish, beef and pork, each of which may harbour the larval stage of certain different species of tapeworm. The beef tapeworm spends its larval stage in cattle and its adult stage in human beings and it is acquired by man only through eating raw or undercooked infected beef. It was relatively unknown in Britain before World War II, but then apparently some tapeworm carriers entered the country among refugees and soldiers returning from the Continent and tropical areas. A small number of infected humans can provide a source of infection for large numbers of cattle. (Fig. 1).

When cattle ingest tapeworm ova, their intestinal juices digest the various protective membranes that enclose the embryo and activate the small six-hooked larva. The larva then penetrates the intestinal wall and having passed via the circulation to various muscles, develops into a cysticercus, a small cyst about 1 cm. in diameter. These cysts appear to have a preference for certain muscles and when cattle are slaughtered in the abattoir, inspection incisions are made in the muscles of mastication, active leg muscles and the heart. If there is evidence of gross infection the meat is condemned, but if the cysts are few in number the beef is stored at 20° F. for three weeks, during which time the cysts will die. Such treated carcasses are downgraded and may only be used for processed meat preparations.

[3] Tomlinson, T. G., and Munden, M. J. *Proc. Inst. Sewage Purification*, 1948 (1), 168.

Meat inspection represents the first line of defence against human tapeworm infections, but it has its limitations. Since most infected cattle have only light infections, it is only a matter of chance whether an inspection incision will uncover a cysticercus. In practice it is quite impossible to mutilate carcasses with numerous incisions to look for widely dispersed cysts and it is understandable that infected meat does occasionally reach the public. It is here, therefore, that the second line of defence, adequate cooking of beef, must operate. Beef

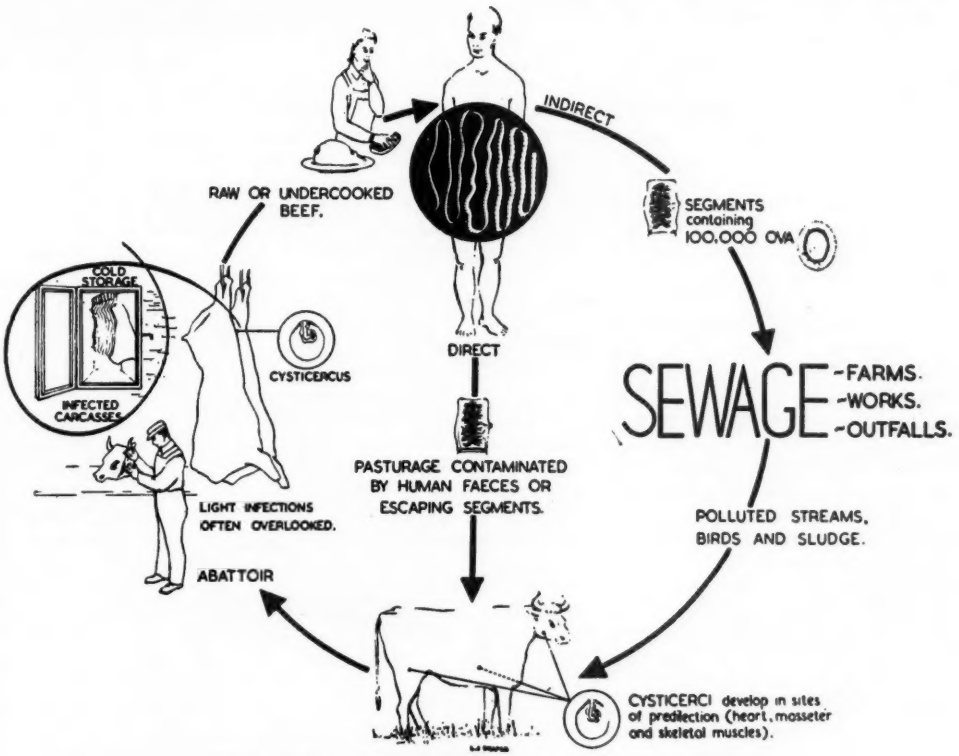


Fig. 1.—Life cycle of the 'beef tapeworm,' *Taenia saginata.*

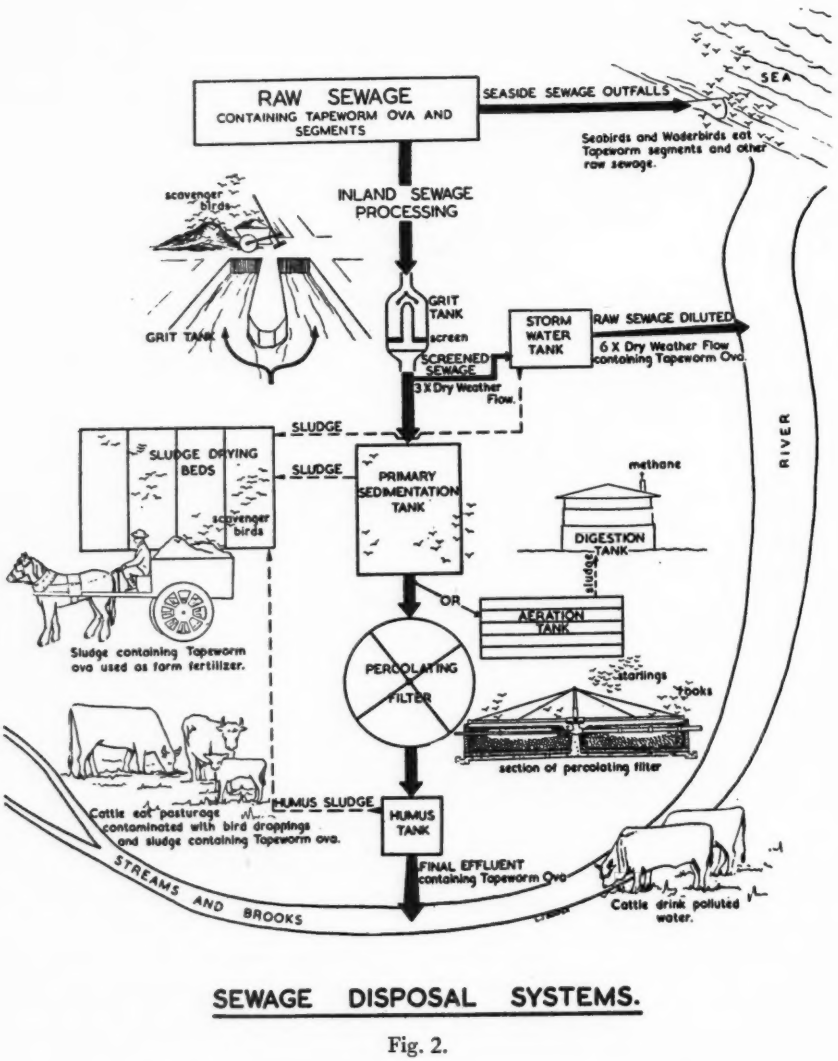
should be well cooked to ensure that meat is heated above the thermal death point of cysticerci, which is about 57° C. (158° F.). There is another method in infection which concerns chiefly women, who are more often infected than men, as it appears they become infected in the kitchen when they taste raw meat or inadvertently eat cysticerci which, being very sticky, may adhere to their hand when preparing food. Once the cysticercus is ingested by man the head invaginates and attaches itself to the intestinal wall. Growth begins and gravid segments are produced in about 60 to 80 days.

Some people may harbour tapeworms for years with no apparent symptoms, while other persons may suffer debilitation and chronic abdominal discomfort. Cysticerci apparently do no harm to cattle, but the lengthy cold storage of carcasses and the downgrading of infected meat results in considerable losses in time and money estimated to range from £100,000 to £500,000 annually.

Tapeworm eggs may come in contact with cattle in several ways. As has been indicated they may be deposited directly on to pastures as a result of indiscriminate defaecation, or segments may escape independently of defaecation by migrating spontaneously through the anal sphincter. Although such direct contact between human and cattle hosts is known to exist, it appears to be a relatively rare occurrence in Britain. If such contacts were common we would expect to find a high percentage of cattle with gross infections of cysticerci as a result of the ingestion of herbage grossly contaminated by deposited segments. Further, cattle which have been found to be infected with cysticerci, and which have been traced back to their farm of origin, can rarely be shown to have been associated with a particular tapeworm carrier. Attempts to find tapeworm

carriers by tracing infected cattle has proved to be a tedious and fruitless task, revealing that infected cattle come from widespread areas with little or no correlation with infected human beings. Indeed, it appears that the majority of individuals infected with tapeworms

The sewage disposal systems in Britain vary from place to place (Fig. 2). The simplest form of disposal is that employed by seaside cities and towns, where raw sewage is drained by gravity into a nearby sea or estuary. Formed faecal matter is emptied into the estuaries



SEWAGE DISPOSAL SYSTEMS.

Fig. 2.

are located in urban areas where they have little or no direct contact with cattle. From the epidemiological standpoint, we may ask how contact between man and beast is made. It seems highly probable that, in Britain, tapeworm eggs pass into sewage systems and negotiate the various processes to be ultimately disseminated over a wide area, where they gain access to grazing cattle.

when one might think that it is put out of harm's way. But it is important to recognise that faecal matter provides a rich source of food for seagulls, wader birds and other scavengers which feed in sewage-infested water. And it has been thought that birds may be able to carry infective organisms from sewage farms and outfalls. We have been able to confirm experimentally that in seagulls, ova of *T.*

saginata can pass through the alimentary tract and retain their viability.

In the processing of sewage inland the coarse faecal material is passed through grids for preliminary screening. These grids, whose spaces vary in width from ¾ inch to 1½ inch, remove much of the debris which finds its way into sewage, as well as quantities of formed faecal matter. The screenings, which are raked away at intervals and stored for burial at a convenient time, may be exposed from one day to two weeks. During this period of exposure scavenger birds have unlimited opportunity to ingest faecal matter, and herein lies another possible method of spread of any ova which may be present.

The screened sewage now passes into either the storm water tank or the primary sedimentation tank, depending upon weather conditions. During the rainy periods, the volume of raw sewage is greatly increased and if three times the normal dry weather flow the excess is stored in the storm water tank until it can be processed during a dry period. After prolonged rain the volume sometimes reaches more than six times the dry weather flow and the raw sewage, now highly diluted, is emptied directly into the river or stream which ordinarily receives only the final effluent.

During normal dry weather operation screened sewage passes into the primary sedimentation tank where tapeworm eggs have been found to settle 18 inches in two hours under ideal conditions. Most primary sedimentation tanks, which are 5 to 12 feet deep, are, however, subject to turbulence due to the constant inflow of sewage and other factors such as the effect of wind. It is doubtful, therefore, whether the normal retention period under the present circumstances of general overloading would permit the sedimentation of a high percentage of tapeworm eggs from primary effluent. It is possible, of course, that some eggs are carried to the bottom by clumping sewage particles, but in any case it is apparent that tapeworm eggs are able to find their way into both primary effluent and sludge.

Primary effluent may be directed into aeration tanks of the activated sludge process, which has no effect on the tapeworm eggs ; or more often to biological filters where it trickles from moving arms on to cinder beds, but the microscopic eggs are able to pass through in the effluent to the humus tank. It can be reasonably presumed that eggs are able to pass through these tanks. There is evidence, however, that sand filtration of humus effluent would remove most of the tapeworm eggs.

The ability of tapeworm eggs to survive in sludge treated in various ways has not been fully ascertained. It appears that the ova are able to resist many months of anaerobic sludge digestion. It is very unlikely, however, that tapeworm eggs will survive in sludge which is subjected to high-temperature treatment processes, such as those used in the recovery of methane. On the other hand tapeworm eggs may survive the more common outdoor drying treatment of sludge for many months. We have kept tapeworm eggs under conditions of high relative humidity for almost a year and they are still in a viable state. The temperature and moisture content of sludge drying beds appear to be favourable to the prolonged survival of such ova.

Hence cattle are being infected with tapeworm eggs from sewage in streams and rivers used by them as a source of water. It is calculated from knowledge of the incidence of tapeworm infection that each gallon of raw sewage contains 1 to 6 tapeworm eggs and it it is estimated 50 per cent. of the cattle in England drink sewage polluted water. The role of birds is difficult to evaluate, but they are able to ingest ova and deposit them in a viable state. Birds picking up morsels out of the tank, filters or sludge, visit pastures on which cattle graze. Seagulls which eat raw sewage in the day usually roost for the night in the local water reservoir. Sludge used for fertiliser may provide a source of infection for cattle. In Germany and South Africa roundworm infections in man have been traced to sludge used as home garden fertiliser.

In conclusion, it is important to control ' beef tapeworm infection ' in Britain, and this may best be instituted at sewage works. Growing communities of people from tropical countries known to be heavily infected with various worms emphasise the need of introducing sand filtration or microstaining of all final effluent. Means and ways must be found of preventing birds from gaining access to raw sewage. Dumping untreated sewage into nearby water must be stopped and replaced by more civilised treatment processes.

Sewer Rat Populations and Their Control

by

Dr. E. W. Bentley

The rat is the only vertebrate that has managed to establish itself in the sewers of Britain. The species concerned is almost always the common rat, *Rattus norvegicus*—though two recent instances of infestations by the ship rat, *R. rattus*, have come to light.

Much effort and money are spent by local authorities on sewer rat control, as they do so much harm.

In the first place, rats are known to move

daily between the sewers and adjacent houses or buildings, often by way of defective drains, and this is undesirable from the public health standpoint. Secondly, heavily infested sewers can and do act as reservoirs for repopulating surface areas which have previously been cleared of rats. Thirdly, rats do considerable damage to the sewer system itself and by their excavations in the earth immediately surrounding faulty sewers may cause quite serious road subsidences. Finally, many rats carry *Leptospira icterohaemorrhagiae*, the organism which causes leptospiral jaundice or Weil's disease. A number of sewermen have died from this.

The sewer itself normally varies from a six-inch earthenware pipe to a 4 feet 9 inch brick barrel. The smaller diameter pipes are quite smooth and rats probably only travel along them for any distance when the flow of sewage is low. Brick barrels are usually ovoid in cross section and each row of bricks at the bottom of the ovoid juts out a little, giving a foothold for rats. The animals keep to the side of the flow unless this is too high. Often they run along in a rather crab-wise fashion with the fore feet a row above the hind feet.

In the sewers it is dark all the time and one might expect the rats to be less nocturnal than they would be on the surface. This seems to be confirmed by the frequency with which rats are seen in sewers during the daytime and from observations made during experiments of the interval which occurs between the laying of baits and their removal by rats.

Activity in many instances appears to depend more on diurnal changes in sewer level. Some sewers, especially near summits of systems may be passable all day and, incidentally, may carry a large rat population because food residues are not washed away as quickly as elsewhere. Other sewers may be running at full bore at peak periods. The best time here for feeding is perhaps immediately after peak flow when tit-bits, such as bread and fish heads, are left stranded on the sides.

Apart from man, sewer rats have no predators to contend with and in this respect are better off than surface-living rats. However, there are several other factors which operate to keep sewer rat populations from unlimited expansion, but they vary in importance according to the sewer system. Where food is scarce, large populations cannot be maintained. This was shown by workers from the Bureau of Animal Population at Oxford in a war-time study of certain heavily-blitzed areas of East London.[1] They found that the proportion of rat-infested manholes was lower in an area where a large proportion of houses had been destroyed than in a similar less damaged district. At the other extreme, sewers which service markets, industrial canteens and restaurants, often carry a high rat population unless control measures are maintained.

Another very important factor is the number of nesting sites available. Rats do not usually lie up in the drain itself but seek out disused pipes (or 'dead eyes') or live in excavations which they make in the neighbourhood of cracks or imperfect joints in the sewer. Where the sewers and drains are in good repair it is difficult for rats to find places to hide and the numbers to be found there are correspondingly reduced.

Another factor which must affect population size is the capacity of the sewer to carry the greatly increased flow which follows heavy rain. In some districts a sudden storm may cause the water level to rise even to street level and one may surmise that in such areas the rat mortality must often be 100 per cent. Even a slight rise in flow level may give rise to a strong current which could be fatal to young rats falling into it.

A further hazard is a variety of chemicals which, apart from methane evolved in the sewer itself, find their way into the sewer from overhead industrial undertakings. Sewermen habitually carry an indicator lamp to warn them of danger but, even so, occasionally have accidents. One imagines that at times, rats are no luckier.

The temperature of the sewers is in general more even than above ground—one of the few favourable features of the environment—and generally warmer during winter and summer than the surface, but cooler than above on very hot days. Evidence on pregnancy rates in sewer rats is still rather scanty but what there is supports the view that breeding continues throughout the year with less fluctuation than among surface-living *R. norvegicus*.

An environmental factor which may influence not so much the final size of the undisturbed sewer rat populations as the speed with which such populations recover where present-day methods of control are practised is the effect of rats on each other. Rats in a particular colony tend to be hostile towards rats from elsewhere. Thus if, for example, three colonies of rats are located along a sewer at points A, B and C, in that order, and have sufficient food and shelter, it is unlikely that the animals in the colony at B will refrain from foraging past or even close up to the rats at A and C. That such discrete colonies do occur in larger sewers has been deduced in three ways: (*a*) it is noticed that a rat when disturbed usually decides to double back past the legs of the observer before it has travelled very far, (*b*) by feeding rats at selected points in the sewer on a bait containing a dye, the colony's range of movement has been de-

[1] Chitty, D. A., and Southern, H. N. (1954). *Control of Rats and Mice*. Oxford: Clarendon Press, 3 vols.

termined by noting the distribution of coloured faecal droppings, (*c*) by depositing known weights of bait at intervals along considerable lengths of the main drain it has been possible to map the feeding distribution of the rats and confirm it by poisoning at suitable points and re-censusing.

The importance of more information on this matter is obvious when it is recalled that existing methods of control depend on poisoning *at the manholes only*. If some rats feed exclusively *between* manholes they will escape poisoning and will represent a breeding nucleus from which a new population can grow.

Accordingly, present research is partly directed towards finding out how frequently and under what circumstances these isolated colonies occur. It may be surmised that it will vary very much from sewer to sewer. Thus it appears likely that where either food or shelter is scarce, all or nearly all the rats in the sewer will pass one or more manholes in their daily foraging and where food and shelter are abundant there will be less movement and, especially where manholes are far apart, control will be more difficult.

Birds on Sewage Farms

by

W. B. Alexander

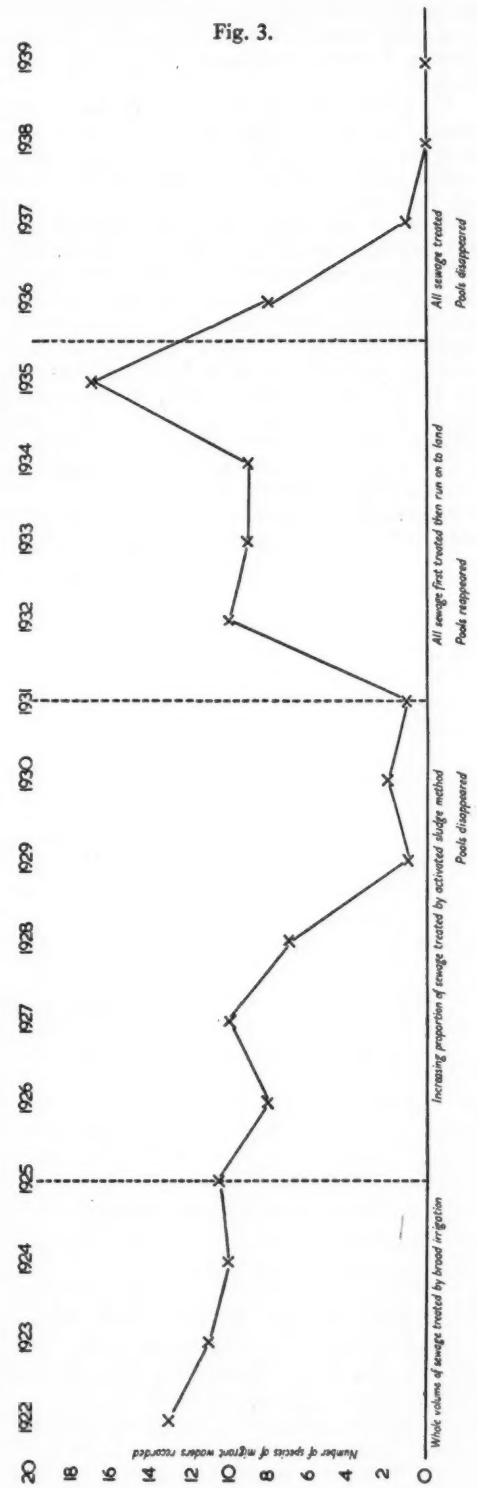
The abundance of birds of a variety of species in many sewage farms, especially those where the sewage is still treated by downward displacement, indicates the rich food supply available. But it appears no scientific study has been made to indicate the actual organisms used as food by any individual species of bird.

The fame of sewage farms among ornithologists is due almost entirely to the fact that during the periods of migration numerous species of wading and aquatic birds, otherwise rarely met with in inland habitats, may be found there, often in considerable numbers. Regular bird-watching at the more attractive sewage farms during the past forty years has added numerous species to the list of many inland counties and has increased our knowledge of migration.

The evidence indicates that water-loving birds pass over all parts of England during migration, commonly at heights too great to be seen from the ground ; that such birds are tempted to descend by the sight of sheets of water ; and that if they find good feeding grounds in the vicinity of such pools they may remain for a few days, or sometimes even for weeks.

Figure 3, based on observations at Reading sewage farm (Manor Farm) from 1922 to 1939, shows the numbers of species of waders

Fig. 3.



observed in each year (not including six species occurring annually, viz. Lapwing, Golden Plover, Curlew, Redshank, Snipe and Jack Snipe).

Up to 1925 the whole volume of sewage was treated by broad irrigation and it will be seen that the number of species of visiting waders varied between 10 and 13. From 1925 to 1931 an increasing proportion of the sewage was treated by the activated sludge method and towards the end of the period pools virtually disappeared and migrant waders were almost absent. From 1931–35 all sewage was first treated in sprinklers and then run on to the land so that pools reappeared and with them waders, the latter reaching a peak of 17 species in 1935. From 1936 onwards all sewage has been completely treated by the activated sludge method, pools disappeared within a year, and since then hardly any migrating waders have visited the farm.

The birds which are attracted by the sprinkler and filters in a modern sewage-treatment works appear to be almost exclusively Starlings, Pied Wagtails and Meadow Pipits.

In discussion, reference was made to the flocks of gulls that feed on the filter beds of some sewage works during the day.[1] They arrive fairly early in the morning and leave as evening approaches for a sheet of water, frequently a reservoir, where they spend the night. Their presence on the reservoir is causing concern, for there is a fear of contamination due to the habit of regurgitation. Attempts to keep them away by various devices, such as ultrasonics or decoys, have not been very successful so far but possibly more precise knowledge of their behaviour is required before they can be diverted. However, although there is evidence suggesting that the area of water occupied during rest has a higher bacterial count than the remainder, the filtration plant can still deal with it successfully.

[1] Spaul, E. A. *J. Instn Wat. Engrs*, 1952, VI, 277.

OBESITY

SECTION I (Physiology) devoted the morning of Friday, September 3, 1954, at the Oxford Meeting of the British Association to a discussion of Obesity. The symposium was introduced by the Chairman, Dr. H. M. Sinclair, Director of the Laboratory of Human Nutrition, University of Oxford. He was followed by Dr. J. M. Tanner, St. Thomas's Hospital Medical School, who presented an account of the classification of differences of body shape in man and their development during growth. Prof. Jean Mayer, Harvard University, then dealt with the normal mechanism of regulation of food intake, and Dr. A. P. Meiklejohn, University of Edinburgh, with the assessment of obesity. Dr. G. C. Kennedy, University of Cambridge, discussed the central nervous control of food intake, and the symposium finished with a contribution from Dr. J. Hammond, C.B.E., F.R.S., University of Cambridge, on the deposition of fat in farm animals.

These papers, with the exception of that by Prof. Jean Mayer, are printed below in the order in which they were delivered.

INTRODUCTION

by

Dr. H. M. Sinclair

WE hear much about undernutrition and malnutrition, by which is usually implied deficiency of vitamins or minerals. In the last few years we have realised that overnutrition, apart from wasting food, is a severe danger to health and life. This has become a subject of particular interest in the United States, for obvious reasons. The danger arises partly from mechanical causes: carrying 20, 30 or even 50 lb. of extra weight, whether fat or trunkloads of baggage, puts more work on the heart, the lungs and the joints; therefore there is an increased incidence in those who are fat of degenerative arthritis of the lumbar spine, hips and knees, of hernias and varicose veins, of accidents presumably because the obese are more clumsy and slower in their movements, and postural changes cause respiratory and circulatory inefficiency with emphysema and diminished vital capacity.

But in the last few years we have begun to realise that the biggest hazard to health is not being overweight but becoming overweight, that the process of growing fat is more dangerous to health and life than is the state of being fat. This can be illustrated by brief reference to atherosclerosis. It is well known that increased blood cholesterol tends to be associated with increased atherosclerosis, and this is not surprising since a large proportion of the atheromatous plaque is cholesterol. But in man increased dietary cholesterol is not in general associated with increased blood cholesterol whereas this is associated with increased dietary fat. It seems probable therefore that excessive consumption of fat will cause on the one hand obesity and on the other atherosclerosis. This may not be the reason for the dramatic increase in coronary thrombosis in recent years in this country; the total consumption of fat in this country now is the same as or very slightly lower than it was pre-war. According to Morris the increase in coronary thrombosis has not been accompanied by any increase in atheroma. I believe the answer may be that the level of total dietary fat is related to atheroma, but that the level of dietary *vegetable* fat is related to clotting upon atheromatous plaques, and hence to coronary thrombosis: the consumption of vegetable fat in this country has risen sharply in recent years (41·4 g. fat/head/day in 1953; 35·1 g. in 1948 and 23·6 g. pre-war).

How severe is the danger from overeating is indicated by the reflection that if you are middle-aged or more, your chances of dying are greater if you are 10 lb. or more

overweight than are your chances of getting cancer of the lung if you are a heavy smoker. Overeating is very much more dangerous than smoking. And not only in the middle-aged, for if you overfeed children they probably mature earlier and die earlier.

An obese person overeats and much has been learnt in the last few years about the control of appetite ; we have two experts to discuss that work—Professor Mayer and Dr. Kennedy. But some people are constitutionally endomorphic—they have more fat-cells than other people ; Dr. Tanner is an expert on that matter. We are, of course, all agreed that when a person gets obese he acquires more fat in already present fat-cells ; the *number* of fat-cells does not increase. We could make him thin by decreasing the number of cells ; we could perhaps cause them to atrophy by feeding a radioactive fat-soluble substance which would be stored in them. Films-stars and ballet dancers could then eat all they wish and still remain slim. But perhaps there would be physiological consequences. Dr. Hammond is, of course, a great authority on the physiology of fat deposition, though in farm animals rather than film-stars.

Finally, if obesity implies waste of food and is dangerous to health and can easily be measured (a subject to be dealt with by Dr. Meiklejohn), ought not the Welfare State to tax it!

Obesity and the Classification of Body Build

by

Dr. J. M. Tanner

It is significant of the times, I think, that a symposium on obesity should begin with a description of the normal varieties of body build in man. Not so very long ago a simple assessment of body weight in relation to height was all that was thought necessary, and ' overweight ' was synonymous with ' obese.' Most obese people are, of course, overweight for their height ; and so are most Olympic weight-lifters and wrestlers. The excess weight is due in the one case to fat, in the other to muscle and bone. Instead then, of equating obesity with ' overweight ' we can equate it with ' overfat,' and methods are now available for estimating the amount of fat in the body, which though neither so convenient nor so accurate as the weighing scales, are perfectly feasible and reasonably valid. This might be thought to be the end of the matter : if a person has over, say, 30 per cent. of his body weight as fat, he is by definition and measurement obese and should be dieted. But the question is, should he ? Mortality statistics (Dublin, 1953) seem to answer unequivocally yes, because he will have a better chance of living longer. However, these are in the main mass statistics and leave unanswered the question as to what happens to the man genetically destined to have a high percentage of fat, when *he* reduces. It may be that no clinical problem exists here (though one would like to be sure about it) and that getting rid of fat is a good thing for everybody, regardless of genetical constitution. Or it may be that the genetically fat man is simply out of luck and either pays on the roundabouts of obesity or loses on the swings of fat reduction. However the clinical situation may be, though, I believe the physiological problem is not so simple that we can equate obesity even with overfatness ; I think we must be prepared to consider obesity in relation to the individual's fundamental body build.

Accordingly, in developing this theme, I intend first to give a brief description of modern classifications of body build ; second to indicate the hereditary background of build, and third, to draw certain analogies between the development of the hereditary potentialities for becoming muscular and the development of similar potentialities for becoming fat.

1. Classifications of Body Build

Amongst the many ways of classifying human physique (see Schreider, 1937 ; Tanner, 1953a, for reviews) there are three that I think should be mentioned here. These are the classification by *somatotypes* (Sheldon, 1940), the dissection by means of *factor analysis* (e.g. Burt, 1949) and the summarisation by relative percentages of the various *tissue components* (e.g. Keys and Brozek, 1953).

Sheldon's system of classifying the human body, called somatotyping, has been in use

now some fifteen years. It is essentially an empirical classification springing from the examination of several thousand normal physiques. Sheldon was primarily responsible for one large theoretical advance in this field : he got rid once and for all of the notion of ' body types '—that is, of discrete and distinct categories into one and only one of which a given individual fits. In their place were substituted *components* of body build, each component being distributed in the population in a similar way to height or weight. By this I mean simply that everyone has some of each component, and most people have about an average amount of each, just as the majority of people are round about average in height.

The number and nature of the components varies from one system of classification to another. Sheldon's somatotyping is founded on three components, called endomorphy, mesomorphy and ectomorphy, and best described by reference to individuals extreme in one component and minimal in the other two. The extreme in endomorphy approaches the spherical as nearly as is humanly possible ; he has a round head, a large fat abdomen predominating over his thorax, and weak, floppy penguin-like arms and legs, with a great deal of fat in the upper arm and thigh, but slender wrists and ankles. Relative to his general size he has also a large liver and spleen and gut, large lungs, and a heart distinguishable in shape from hearts of the extreme mesomorphs and ectomorphs. He has a great deal of subcutaneous fat and in lay parlance might simply be called a fat man, but I am particularly concerned to get across Sheldon's idea that there is more to it than that. If Sheldon is right, fatness is only one of the endomorph's characteristics ; the extreme endomorph's whole body differs in shape from those of the other extremes ; his thoracic and pelvic skeleton is relatively greater in the anteroposterior direction for example, his fingers are shorter, his head rounder. Starve a person high in endomorphy, and he does not turn into something else ; emptying his fat cells does not affect the fundamental structure of his body either in its cell composition or in its spatial extension—or capacity for extension.

The extreme in mesomorphy is the classical Hercules ; in him bone and muscle predominate. He has a cubical massive head, broad shoulders and chest, and heavily muscled legs and arms, with the distal segments strong in relation to the proximal. Relative to his size his heart muscle is also large. He has a minimal amount of subcutaneous fat and the anteroposterior diameters of his body are relatively small.

The extreme in ectomorphy is the linear man ; he has a thin peaked face with a receding chin and high forehead, a thin narrow chest and abdomen, a narrow heart and spindly arms and legs. He has neither much muscle nor much subcutaneous fat, but, relative to his size, he has a large skin area and a large nervous system.

Naturally very few people resemble these extremes ; the majority have a moderate amount of each of the three components. Each component is rated or measured on a 7-point scale so that the majority have as their somatotype, as the set of 3 ratings is called, such numbers as 433, 344 or 352. One difficulty of this classification is that the three components are not independent of each other, which makes for complexity. There are other difficulties too, that we have no time to go into here ; but in the main the classification works, and despite the apparent complexity of the descriptions, properly-trained observers agree very closely on an individual's somatotype, the discrepancy being seldom more than a half rating on the 7-point scale (Tanner, 1954).

We have no time, unfortunately, to discuss the classification of physique by *factor analysis*, which I think holds considerable promise particularly in relation to genetics and experimental embryology ; it has little practical application, however, at present.

Thirdly, there is the classification by *tissue components*. This is complementary to and in no way exclusive of the classifications by somatotype and factor analysis. Tissue component analysis primarily sets out to tell us how much of each tissue a person has in his whole body ; how much fat, how much muscle, how much bone. At present only the fat percentage can be accurately assessed, but we may reasonably

hope that before long the relative amounts of muscle and bone will also be amenable to analysis. This type of analysis can be particularised to tell us something about the spatial distribution of the tissue components, for example as to whether a large muscle percentage is distributed along a tall thin man or across a broad short one, but it does not primarily set out to do this, as the other two classificatory systems do. To my mind the possibilities of soft-tissue roentgenography (Tanner, 1953c) combined with somatotyping have scarcely begun to be realised: just to mention two intriguing findings, both essentially due to Reynolds: (a) the relative amounts of fat, muscle and bone in the calf are practically (and astonishingly) independent of each other (Reynolds and Asakawa, 1950) and (b) the sex of the person X-rayed can be told with 95 per cent. success from the relative amounts of fat and bone in the calf (Tanner, 1953d).

2. Genesis of Body Build

Studies of monozygotic and dizygotic twins, and of monozygotic twins reared in different environments, have shown that about 90 per cent. of the differences between people in body length measurements are due to heredity, assuming reasonable conditions of nutrition during childhood (see Tanner, 1953b). For body breadths the influence of heredity is slightly less, but not much so. Table I shows the correlation coefficients for height, weight and head breadth and length for various sorts of twins. It can be seen that the difference between identicals reared together and reared apart is statistically insignificant, and that the identical/non-identical difference is practically as large for weight as for the bony measurements. In other words, body weight is fundamentally and to a very great degree controlled by hereditary factors. There are no figures, so far as I know, on the heritability of body fat percentage: doubtless the twin correlations would be somewhat lower because non-hereditary factors such as dieting, exercise and worry certainly affect the laying down of fat in the cells more than they affect any other aspect of physique. But I am sure that even for fat percentage the hereditary factors under normal circumstances would still be very substantial.

Table I

Correlation coefficients of body measurements between pairs of identical and non-identical twins and siblings. From *Clinical Genetics*, ed. A. Sorsby. London: Butterworth, 1954.

Dimension	Identical reared together	Identical reared apart	Non-identical	Sibs.
Height	0·932	0·969	0·645	0·600
Weight	0·917	0·886	0·631	0·584
Head breadth	0·908	0·880	0·654	
Head length	0·910	0·917	0·691	

Two measurements of the same person usually have a correlation of 0·99 for height and weight and 0·97 for head breadth and length.

3. Endomorphy versus Obesity

Lastly I want to discuss a question which seems to me the point of engagement between the disciplines of anthropology, nutrition and physiology, and which might be headlined as 'endomorphy versus obesity.' To start doing this let me draw certain analogies between developing the hereditary potentialities for becoming fat, and developing those for becoming muscular.

People high in mesomorphy—the muscular people—must either have more muscle cells than people low in mesomorphy or else have bigger cells. There is no certain answer yet as to which of these statements is true, but the odds are very strongly in favour of the former, that is, that high mesomorphs have more muscle fibres. Now the differentiation of muscle fibres is complete in man by about the seventh intrauterine month; after this time all that happens is that the fibres enlarge and the amount of actomyosin and other constituents in each increase. This increase is steady throughout childhood, with an extra fast spurt at adolescence, particularly in boys. Very likely there is a decrease in middle and old age, though there are no critical studies of that. But even in young adults the muscle fibres do not reach their maximum possible size (at least in our culture), unless special steps are taken to develop them. This can be done by exercising with heavy

weights three or four times a week or in some other way. The muscles under these circumstances enlarge, some more than others, and probably stay at the increased size so long as the training is maintained ; when it ceases they go back to their former size (Tanner, 1952). Animal work provides strong evidence that the training increase is solely in fibre size and not in fibre number.

Now you will see where this line of thought is leading us. I wish to indicate the analogy between muscular training and obesity. First of all, about people high in endomorphy ; do they have more fat cells than those low in it ? I suspect they do, and with my surgical colleagues I am at present trying to get some definite evidence on this point, but, of course, I must make it perfectly clear that what follows is speculation or perhaps, less opprobriously, hypothesis only. If high endomorphs do have more fat cells, they presumably acquired them in intrauterine life, too, and the problem of endomorphy is a problem of embryology—why is it that people with large anteroposterior skeletal diameters and other physical attributes also develop large numbers of fat cells ? So far as I am aware there are no even tentative answers to this at present.

To account for the alterations in body fat during growth we must postulate a slightly more complicated system of occurrences than for muscle fibre enlargement. Body fat increases from birth to about 6 months, then decreases to about 5 years, stays stationary for a year or two, then gradually increases from that time onwards, except for a check at adolescence. As the individual gets older the fat cells fill up, under ordinary circumstances. (We cannot be sure they do not increase in number, but I am assuming this to be improbable.) If the individual has few fat cells—that is if he is (according to my hypothesis) low in endomorphy—however much the cells fill up he will not become ' obese ' by the 30 per cent. of body weight criterion (though to the acuter family eye he may begin to look a bit bloated). On the other hand, if he has large numbers of fat cells—is high in endomorphy—even a relatively small increase of fat in each cell will make him clinically obese.

I want to put forward the view that we need, for investigative purposes anyway, to distinguish these two situations. We need a term for the amount of fat present per fat cell. We cannot define obesity in this way any more than we would define muscularity by the amount of actomyosin per muscle fibre, but I suggest the term *lipoplethory*, following Lawrence (1951), could be used to designate this relative fulness of cells with fat. We could then take the view that a person of any build might become, *a priori*, lipoplethoric, but that only those relatively high in endomorphy could produce a sufficiently high percentage of body fat to be called obese.

On this view, just as training increases the amount of actomyosin in the muscle fibre, so gluttony increases the degree of lipoplethory. Nobody knows what makes the muscles enlarge under training, and nobody knows why fat cells enlarge under gluttony. But clear separation of the concepts of endomorphy, lipoplethory, and obesity may make thinking in this field somewhat easier and physiological relationships somewhat clearer. Let me take as an example the relationship between body fat and serum cholesterol concentration (Tanner, 1951). The relation, though statistically certain, is not a very close one as measured. But is seems to me likely that the true relation is between cholesterol and lipoplethory, not the amount of fat as such. And if this is really the case the connection might be very much closer than it appears at first sight.

From the point of view of therapeutic medicine probably the percentage of weight carried as fat is what matters. From the standpoint of preventive medicine, however, it may well be that a guide to the number of fat cells an individual possesses would be the most valuable thing. In well-nourished countries people with many fat cells are the ones who need to diet to prevent themselves becoming obese, and prevention is better than cure. From this approach, and also from the approach of physiological research I think the distinction of and the investigation into lipoplethory may have considerable importance.

REFERENCES

BURT, C. (1949) : ' Subdivided factors.' *Brit. J. Psychol. (Stat. Sec.)* **2**, 41–63.

DUBLIN (1953) : Cited in Sinclair (1953) : ' Assessment and results of obesity. *Brit. med. J.* **2**, 1404–1407.

KEYS, A., and BROZEK, J. (1953) : ' Body fat in adult man.' *Physiol. Rev.* **33**, 245–325.
LAWRENCE, R. D. (1951) : ' Types of diabetes.' *Brit. med. J.* **1**, 373–375.
REYNOLDS, E. L., and ASAKAWA, T. (1950) : ' A comparison of certain aspects of body structure and body shape in 200 adults.' *Amer. J. phys. Anthrop., N.S.* **8**, 343–365.
SCHREIDER, E. (1937) : *Les types humains.* Paris : Hermann.
SHELDON, W. H. (1940) : *The varieties of human physique.* New York : Harper.
TANNER, J. M. (1951) : ' The relation between serum cholesterol and physique in healthy young men. *J. Physiol.* **115**, 371–390.
TANNER, J. M. (1952) : ' The effect of weight-training on physique. *Amer. J. phys. Anthrop., N.S.* **10**, 427–462.
TANNER, J. M. (1953a) : ' Growth and constitution in Kroeber,' A. L. ed. *Anthropology Today* : Univ. Chicago Press.
TANNER, J. M. (1953b) : ' Inheritance of morphological and physiological traits.' In Sorsby, A. ed. *Clinical Genetics.* London : Butterworth.
TANNER, J. M. (1953c) : ' Growth of the human at the time of adolescence.' *Lect. Sci. Basis Med.* **1**, 308–363.
TANNER, J. M. (1953d) : ' Investigations on the growth and physique of the human.' Ph.D. Thesis, London University.
TANNER, J. M. (1954) : ' Reliability of anthroposcopic somatotyping.' *Amer. J. phys. Anthrop., N.S.* **12**, 257–265.

The Assessment of Obesity

by

Dr. A. P. Meiklejohn

One of the most useful records about his patients that a doctor can keep is their weight, measured at regular intervals. A sudden and unexplained loss of weight may be the first sign of many diseases and often occurs at a curable stage. In the opposite direction, any excessive gain in weight has important implications for the future health of the patient. The experience of life insurance companies has shown that grossly overweight people are poor risks. The reason for this is that they are prone to a variety of misfortunes—mechanical and metabolic—which I shall discuss shortly. It is, therefore, pertinent to study the nature of the material that accumulates in the body when a person puts on weight, and to ask in what way this material is related to, or responsible for, the complications of obesity. In other words, when we stand on the bathroom scales and note that our weight is up, what are we trying to measure ?

Body Weight as an Index of Body Fat

The easy answer is : ' fat.' But what precisely do we mean by that ? So much pure chemical lipid, like a packet of margarine. Certainly, for research purposes it may be useful and interesting to measure the total amount of chemical fat in the body and this cannot be assessed with any accuracy from the body weight. As Dr. Tanner has already indicated, there are great variations in body build and skeletal size ; although endomorphs may carry more fat than mesomorphs of the same height, this is not necessarily shown in their weight. People may gain weight from putting on materials other than fat ; athletes in training may put on muscle, patients with kidney failure, water.

The methods at present available for estimating total chemical fat in the living body are complicated ; we can arrive at an estimate by determining the specific gravity of the body by the method of Archimedes, as first used by Behnke (1941) and developed by Keys in the U.S.A. But the weighing of middle-aged people under water is scarcely a method for routine clinical use. Alternatively the method developed by McCance (1951) in this country can be used ; the total body water is measured by the dilution of an introduced chemical substance which is freely diffusible in all the body fluids, and from this (with some assumptions) the ' lean body mass ' is calculated. This, deducted from the total body weight, leaves the weight of fat and bones ; the latter can be estimated from the weight of a comparable dry skeleton.

Measurement of ' Adipose Tissue '

It is, however, not the amount of chemical fat in the body that we really want to measure. Keys and Brozek (1953) have pointed out that when a healthy person puts on weight, this is not just a simple matter of depositing chemical lipids. They suggest that it is due to the deposition of ' adipose tissue ' which is composed not only of fat, but also of ' cellular material ' (containing protein) and water. They further suggest that this adipose tissue has a relatively constant composition, with the same proportions of fat, protein

and water in different people. Dr. Passmore, Miss Thow and I (1954) have lately obtained evidence that makes us doubt this. We fed three thin but healthy young men about 1,600 calories of extra food daily and found that they put on weight, but not as much as they should have done had they laid down ' adipose tissue ' as defined by Keys and Brozek. All their gain in weight could be accounted for in terms of chemical substances : protein and lipids. There was no evidence of an associated retention of water. I shall return to this interesting result later on.

Meantime, to assess the amount of adipose tissue in the body, probably the best method is to measure the thickness of skinfolds. The cells in the body capable of taking up variable amounts of fat are limited in number and confined to certain areas (Edwards, 1950). About half of them lie immediately under the skin, so by measuring the thickness of the subcutaneous tissues we have a fair estimate of their bulk. Many people have pointed out that the measurement of skinfolds is a good way of estimating adipose tissue in the body. I began to measure them myself about twelve years ago in the course of some nutrition surveys, in the hope that they might prove a useful index of nutritional state. Then, as now, I concluded that such measurements tell no more about the immediate health of a patient than can be learned from looking at him and weighing him.

What is Obesity ?

Let me return to my original question : what are we trying to measure ? We want to measure obesity—a word which, despite its earlier meanings, has for some centuries meant ' corpulence.' We want a reliable means of estimating corpulence, because we want to be able to recognise those patients who are liable to suffer later from its recognised complications.

Complications of Obesity

It is worthwhile briefly to consider these complications.

1. *Mechanical.* First, there are a number of complications which are the mechanical result of excessive weight, backache, arthritis of the knees and hips, flat foot, ruptures, varicose veins, shallow respiration leading to chronic bronchitis.

2. *Metabolic.*

   (*a*) *Diabetes mellitus.* In the small proportion of obese people who become diabetic, the reduced tolerance for sugar appears to be related to the amount of lipid in their tissues, because if this fat is removed by underfeeding them, their sugar tolerance almost always improves. We do not yet know the complete explanation of this. The answer is likely to be found in a more precise study of the intermediary metabolism of lipids and carbohydrates in man. In the course of such a study it might become useful to be able to measure the total body lipid.

   (*b*) *Hypertension.* Some obese people have high blood pressure. Fortunately it is usually of a benign kind. In my experience it seldom improves with weight reduction and although it may have a metabolic origin, I do not think that it can be related directly to the amount of lipids or adipose tissue in the body.

   (*c*) *Disturbances of Cholesterol Metabolism.* Obese people develop gall-stones (made of cholesterol) more frequently than others. More important, they get deposits of cholesterol (atheroma) in their arteries. The cause of this is the prime geriatric problem of the present age. Ancel Keys and his colleagues have produced impressive evidence that people who live on diets rich in fat have raised levels of cholesterol in the blood. Moreover, people who develop coronary thrombosis as a result of atheroma, are also often found to have an elevated blood cholesterol level (e.g. Oliver and Boyd, 1953). But it is not a logical inference to conclude that therefore atheroma in the coronary arteries and other vessels is due simply to eating too much fat.

KEYS, A., and BROZEK, J. (1953) : 'Body fat in adult man.' *Physiol. Rev.* **33**, 245–325.
LAWRENCE, R. D. (1951) : 'Types of diabetes.' *Brit. med. J.* **1**, 373–375.
REYNOLDS, E. L., and ASAKAWA, T. (1950) : 'A comparison of certain aspects of body structure and body shape in 200 adults.' *Amer. J. phys. Anthrop., N.S.* **8**, 343–365.
SCHREIDER, E. (1937) : *Les types humains*. Paris : Hermann.
SHELDON, W. H. (1940) : *The varieties of human physique*. New York : Harper.
TANNER, J. M. (1951) : 'The relation between serum cholesterol and physique in healthy young men. *J. Physiol.* **115**, 371–390.
TANNER, J. M. (1952) : 'The effect of weight-training on physique. *Amer. J. phys. Anthrop., N.S.* **10**, 427–462.
TANNER, J. M. (1953a) : 'Growth and constitution in Kroeber,' A. L. ed. *Anthropology Today* : Univ. Chicago Press.
TANNER, J. M. (1953b) : 'Inheritance of morphological and physiological traits.' In Sorsby, A. ed. *Clinical Genetics*. London : Butterworth.
TANNER, J. M. (1953c) : 'Growth of the human at the time of adolescence.' *Lect. Sci. Basis Med.* **1**, 308–363.
TANNER, J. M. (1953d) : 'Investigations on the growth and physique of the human.' Ph.D. Thesis, London University.
TANNER, J. M. (1954) : 'Reliability of anthroposcopic somatotyping.' *Amer. J. phys. Anthrop., N.S.* **12**, 257–265.

THE ASSESSMENT OF OBESITY

by

Dr. A. P. Meiklejohn

ONE of the most useful records about his patients that a doctor can keep is their weight, measured at regular intervals. A sudden and unexplained loss of weight may be the first sign of many diseases and often occurs at a curable stage. In the opposite direction, any excessive gain in weight has important implications for the future health of the patient. The experience of life insurance companies has shown that grossly overweight people are poor risks. The reason for this is that they are prone to a variety of misfortunes—mechanical and metabolic—which I shall discuss shortly. It is, therefore, pertinent to study the nature of the material that accumulates in the body when a person puts on weight, and to ask in what way this material is related to, or responsible for, the complications of obesity. In other words, when we stand on the bathroom scales and note that our weight is up, what are we trying to measure?

BODY WEIGHT AS AN INDEX OF BODY FAT

The easy answer is : 'fat.' But what precisely do we mean by that? So much pure chemical lipid, like a packet of margarine. Certainly, for research purposes it may be useful and interesting to measure the total amount of chemical fat in the body and this cannot be assessed with any accuracy from the body weight. As Dr. Tanner has already indicated, there are great variations in body build and skeletal size ; although endomorphs may carry more fat than mesomorphs of the same height, this is not necessarily shown in their weight. People may gain weight from putting on materials other than fat ; athletes in training may put on muscle, patients with kidney failure, water.

The methods at present available for estimating total chemical fat in the living body are complicated ; we can arrive at an estimate by determining the specific gravity of the body by the method of Archimedes, as first used by Behnke (1941) and developed by Keys in the U.S.A. But the weighing of middle-aged people under water is scarcely a method for routine clinical use. Alternatively the method developed by McCance (1951) in this country can be used ; the total body water is measured by the dilution of an introduced chemical substance which is freely diffusible in all the body fluids, and from this (with some assumptions) the 'lean body mass' is calculated. This, deducted from the total body weight, leaves the weight of fat and bones ; the latter can be estimated from the weight of a comparable dry skeleton.

MEASUREMENT OF 'ADIPOSE TISSUE'

It is, however, not the amount of chemical fat in the body that we really want to measure. Keys and Brozek (1953) have pointed out that when a healthy person puts on weight, this is not just a simple matter of depositing chemical lipids. They suggest that it is due to the deposition of 'adipose tissue' which is composed not only of fat, but also of 'cellular material' (containing protein) and water. They further suggest that this adipose tissue has a relatively constant composition, with the same proportions of fat, protein

and water in different people. Dr. Passmore, Miss Thow and I (1954) have lately obtained evidence that makes us doubt this. We fed three thin but healthy young men about 1,600 calories of extra food daily and found that they put on weight, but not as much as they should have done had they laid down 'adipose tissue' as defined by Keys and Brozek. All their gain in weight could be accounted for in terms of chemical substances: protein and lipids. There was no evidence of an associated retention of water. I shall return to this interesting result later on.

Meantime, to assess the amount of adipose tissue in the body, probably the best method is to measure the thickness of skinfolds. The cells in the body capable of taking up variable amounts of fat are limited in number and confined to certain areas (Edwards, 1950). About half of them lie immediately under the skin, so by measuring the thickness of the subcutaneous tissues we have a fair estimate of their bulk. Many people have pointed out that the measurement of skinfolds is a good way of estimating adipose tissue in the body. I began to measure them myself about twelve years ago in the course of some nutrition surveys, in the hope that they might prove a useful index of nutritional state. Then, as now, I concluded that such measurements tell no more about the immediate health of a patient than can be learned from looking at him and weighing him.

What is Obesity?

Let me return to my original question: what are we trying to measure? We want to measure obesity—a word which, despite its earlier meanings, has for some centuries meant 'corpulence.' We want a reliable means of estimating corpulence, because we want to be able to recognise those patients who are liable to suffer later from its recognised complications.

Complications of Obesity

It is worthwhile briefly to consider these complications.

1. *Mechanical.* First, there are a number of complications which are the mechanical result of excessive weight, backache, arthritis of the knees and hips, flat foot, ruptures, varicose veins, shallow respiration leading to chronic bronchitis.

2. *Metabolic.*

   (*a*) *Diabetes mellitus.* In the small proportion of obese people who become diabetic, the reduced tolerance for sugar appears to be related to the amount of lipid in their tissues, because if this fat is removed by underfeeding them, their sugar tolerance almost always improves. We do not yet know the complete explanation of this. The answer is likely to be found in a more precise study of the intermediary metabolism of lipids and carbohydrates in man. In the course of such a study it might become useful to be able to measure the total body lipid.

   (*b*) *Hypertension.* Some obese people have high blood pressure. Fortunately it is usually of a benign kind. In my experience it seldom improves with weight reduction and although it may have a metabolic origin, I do not think that it can be related directly to the amount of lipids or adipose tissue in the body.

   (*c*) *Disturbances of Cholesterol Metabolism.* Obese people develop gall-stones (made of cholesterol) more frequently than others. More important, they get deposits of cholesterol (atheroma) in their arteries. The cause of this is the prime geriatric problem of the present age. Ancel Keys and his colleagues have produced impressive evidence that people who live on diets rich in fat have raised levels of cholesterol in the blood. Moreover, people who develop coronary thrombosis as a result of atheroma, are also often found to have an elevated blood cholesterol level (e.g. Oliver and Boyd, 1953). But it is not a logical inference to conclude that therefore atheroma in the coronary arteries and other vessels is due simply to eating too much fat.

Many other factors besides dietary fat may influence the level of cholesterol in the blood and its deposition in the walls of the arteries. We should not be too impressed by the fact that the blood cholesterol level can be reduced temporarily by feeding a fat-free diet; this is quite probably due to the lack of fatty acids in the intestine which are necessary to enable the body to reabsorb the 1 g. or so of cholesterol which the body normally excretes in its own bile.

(*d*) *Comment.* Many things in the body change with overfeeding. For one, several endocrine (ductless gland) secretions are affected. Our own evidence is that with overfeeding, the cells of the body may undergo a concentration of their materials—not only lipids but also protein. The male sex hormones when injected, cause a retention of protein in the body, so perhaps they are concerned in this process. In which case we might have the explanation of why obesity is more common in women than in men. Clearly we have a great deal more to learn about the reactions of the body to overfeeding and its relation, direct or indirect, to arterial degeneration.

The Assessment of Obesity

Pending the results of this fundamental research, how is a doctor to tell whether his patient is obese? I suggest that his best guide is to look at him. If he wants a factual record of what he sees he can, like Dr. Tanner, photograph him. A simpler method is the old-fashioned one of weighing the patient. I still think that change in the weight of the body—no matter whether it is lipid, protein or water that affects it—is a useful yardstick for measuring the effects of overfeeding and for studying the natural history of the complications that may attend it.

Standards of Body Weight

If this is agreed, then obviously we need reliable standards for judging whether people are overweight for their race, sex, age and height. At first it might seem that the best yardstick would be the average weight of comparable but apparently healthy people. But even were such averages available for the present population of Britain (which they are not) it is doubtful if they would provide a valid index of normality; there are too many obese people in our population. Tables of 'standard' or 'ideal' weights are published in many textbooks of medicine and dietetics on both sides of the Atlantic. The origin of these tables is interesting. They are based on the average heights and weights of many thousands of Americans seeking life insurance between the years 1885 and 1900. They were weighed without overcoats, but in the street clothing of that time, whalebone and all. The figures so collected were first published in the Medico-Actuarial Investigation of 1912. Clinics and life insurance companies have used them ever since, apparently with satisfaction, except that a departure of about $\pm$ 10 per cent. from the average figures came to be recognised generally as unimportant. In fact, patients with diabetes keep in the best health if they are 10 per cent. below these standards. This might also be true of people without diabetes.

An apparent departure from these accepted standards were the new tables brought out by Metropolitan Life Insurance Co. (U.S.A.) in 1942–43; these have since been adopted in several textbooks. But these new tables seem to be based on the same old data, with some theoretical adjustments; the tall people have been thinned a bit, the short people fattened, the average heights have been slightly reduced and the curve connecting the points for each age and height smoothed out. Thereafter a spread of figures of about one standard deviation on each side of the average has been calculated. The range of figures so obtained is then divided up into three arbitrary, overlapping columns suitable for people of 'small, medium and large frame' respectively. Finally, a single age standard was set for people 'aged 25 and over,' in the belief that it is unwise for anyone ever to exceed their weight at that age. In fairness, it must be said that these tables were intended

primarily for health education purposes; they were never meant to be a medical definition of 'ideal' weights. There are, however, obvious drawbacks to their use in Britain:

1. They are based originally on the stature of American people.
2. They offer no means of judging whether a patient has a 'small, medium or large frame' and so of deciding in which category he should be placed.
3. Is it really true that it is unwise for anyone ever to exceed their weight at 25? Certainly the majority of people do, and I know of no evidence that their health is impaired thereby, provided that they do not become obese. Have the life insurance companies any evidence that a small gain in weight throughout middle life is anything but normal? If they have, they should publish it.
4. These tables become even more artificial when they are further simplified in the way adopted by a daily newspaper a few weeks ago in an article on slimming. The range of figures under each heading 'small, medium and large frame' were simply averaged to provide three single standard figures for each height.

I think it fair to say that we could do with some fresh standards that will apply more directly to the population in Britain to-day. The best figures at present available are those collected by the Ministry of Food in 1943 for the heights and weights of several thousand housewives and workers in small industries. A useful analysis of these figures has been given by Kemsley (1950). But he specifically warns against using them as standards of height and weight. Personally I think he is rather too modest about the value of his work. His figures might make a useful interim standard for use in Britain. Although the nutritional health of the population in 1943 may not have been ideal, at least we knew then more precisely than at any other time what the British people were eating. It is interesting to compare Kemsley's figures with the original American standards. It is evident that British people in 1943 had nothing like the same tendency to obesity in later life as had prosperous Americans in 1885-1900. Harries and Hollingsworth (1953) have published evidence that the average weight of adults has increased since that time.

Need for Research on Anthropometric Standards

It would obviously be useful to know the average dimensions of the present adult British population. It seems amazing that there is no satisfactory information on this, despite repeated pleas for a proper anthropometric survey.

The British Association itself organised and carried out the first thorough anthropometric survey in Britain. The results were published in 1883. Could we not do so again! The last survey was undertaken primarily in an attempt to define different racial groups within the British Isles, but brought to light the fact that stature was related then to economic circumstances rather than to racial origin. There has since been a revolution in standards of living and consequently of nutrition. A new survey of British stature would have to be related more to the effects of overfeeding than of underfeeding. It would be useful if it could be integrated with a survey of the complications of obesity. Here is a fine field for research workers in gerontology and perhaps the new College of General Practice could assist. The Medical Research Council has a Committee on Growth and Form that could advise.

We ought to be able to give a measure of 'ideal' weight for British people according to racial origin, age, sex, height and somatotype. But much more research is needed before we can do this and meantime do we really know what we are trying to measure when we stand on the bathroom scales?

The Central Nervous Control of Food Intake

by

Dr. G. C. Kennedy.

All obesity is caused by overeating. This is a statement of fact and not an attempt to explain the complex aetiology of obesity.

In the past it has allowed us to reject theories about obesity which were thermodynamic nonsense. Now, it helps us to recognise over-simplifications in current theories of appetite. Some species rarely become obese. It follows thermodynamically that the final determinant of appetite in them must be the calorie content of their food. But whatever governs the amount a pig eats, it is not calories. Man exhibits an intermediate degree of control of his fat content. If obesity has many causes, then many factors must influence appetite, and we must avoid theories which depend upon the recognition of only one quality in food.

Unicellular organisms ingest food indiscriminately, and seem unable to modify their feeding behaviour in response to changes in nutrition. Therefore, if they encounter a relative excess of energy foods over protein, they become fat. They are then unable to divide, so eventually they die. In the higher organisms obesity is associated with a diminished expectation of life. So the maintenance of a reasonably constant fat content by some animals has come to be regarded as an example of homeostasis. It depends upon the possession of a central nervous system to control feeding behaviour, and is seen in organisms as simple as hydra. In higher forms the modification of feeding behaviour in relation to changes in bodily state may be very complex. The active animal becomes hungry, the lazy or caged one has less appetite ; conversely the hungry animal is restless and active and the gorged one sleeps. Sexual activity may affect appetite; for example, the rat eats less during oestrus than it does during dioestrus. During lactation the food intake may be trebled. Rats also increase their food intake greatly in low environmental temperatures. And in most animals thirst inhibits appetite. As lesions in various parts of the hypothalamus may affect food intake, water balance, body temperature, spontaneous activity and sex behaviour either separately or together, this part of the central nervous system is clearly of great importance to the control of the energy balance of the body.

The hypothalamic influence upon appetite has been studied most in the rat, probably because the hypothalamus dominates the feeding behaviour of this species, and its partial destruction is followed by considerable changes in behaviour. In general, the action of hypothalamic centres on the body is automatic, reflex or vegetative in character. This is probably the explanation of their importance to the appetite of the rat. Such small animals eat an enormous amount of food in relation to their weight, to compensate for the loss of heat from their relatively large surface. The time of survival of the rat denied food is, therefore, very short, and it is clearly of importance that it should have a powerful, automatic drive to eat energy foodstuffs. Also as we have already seen, animals which maintain their fat stores relatively constant from day to day, as the rat does, and do not become obese, must determine their satiety level by the calorie content of the food they eat. The areas in the hypothalamus of the rat which affect appetite have been very accurately localised by Brobeck and his co-workers. In a recent definitive account of their work (Anand and Brobeck, 1951) they described two positions in which lesions produced the maximum effect, a medial one near the ventromedial nuclei, and a second area about a millimetre lateral to this. Destruction of the medial area caused a period of increased food intake, or hyperphagia as Brobeck called it, which lasted until the animal became fat. The food intake then returned to more normal levels, but hyperphagia could be induced again by restricting the diet for a time to reduce the weight. Destruction of the second, lateral hypothalamic area abolished eating altogether and the rat died of starvation. Here we have evidence of the existence of two areas, one concerned with the control of satiety, the other with the initiation of feeding behaviour, the balanced activity of which might explain most of the observed facts about the normal rat's feeding.

So far we have considered only the thin rat. But what happens when the rat gets fat ! Can anything be shown to affect its food intake, or is it simply voracious and omnivorous as some of the early reports on experimental obesity suggested ? Adolph (1947) fed normal rats on a variety of unpalatable or bulky diets in order to

demonstrate that bulk and flavour had little influence on the rat's appetite. He concluded that the rat ate for calories. But it is important, I think, to realise that the rat is perfectly well able to appreciate the qualities in Adolph's diets which led him to regard them as unpleasant. I have offered such diets to rats with a choice of alternatives, and they unhesitatingly rejected the unpleasant or bulky diets. When I gave these diets to hypothalamic fat rats, I found that there was no change of preference, but that the fat rat, unlike the thin one, appeared to be able to show its preference by refusal of the unpleasant diets even when no alternative food was given. But it could do this only so long as it had fat to live on. As it lost weight, its behaviour gradually returned to that of the thin rat. This increase in discrimination against unpleasant diets is shown also by old, naturally fat rats, and by mice made obese by gold thioglucose, so it is probably an attribute of obese animals in general. The fat animal seems to have much more of the gourmet about it than the thin one. I have suggested that this discriminative behaviour is a release phenomenon characteristic of animals in which the hypothalamic factor in appetite is naturally weak or has been damaged by operation. (Kennedy, 1950, 1951).

It follows from what has been said, I think, that the hypothalamic centres must be sensitive to some quality of the food related to its calorie content. There are broadly three ways in which this could be achieved, and each has been proposed by different workers. Brobeck believes that the calorimetry is direct, that is, the hypothalamus is sensitive to the heat released during the metabolism of food. Mayer believes the stimulus to be variations in blood sugar. I believe that neither of these theories fits all the facts. If the animal is to remain in energy balance for long periods, it must have some sort of energy memory. I have suggested that this is provided by direct sensitivity to the state of the fat depots (Kennedy, 1953). This could be achieved through the effect of fat turnover on blood metabolites, but there are other possibilities, such as a direct nervous effect through the sympathetic innervation of the fat cells. Mayer has established a convincing case for blood glucose being one of the stimuli to which the central nervous system is sensitive, but I believe it is highly probable that there are others.

The voracious feeding of the hyperphagic rat is highly characteristic, but is not invariable. In milder degrees of obesity, or even in severe obesity of a slightly different type due to large lesions (Hetherington and Ranson, 1939) there may be no marked hyperphagia. Such differences in behaviour may account for discrepancies in the literature, such as the report by Miller, Bailey and Stevenson (1950) that ' hunger drive ' is reduced in hypothalamic obesity. So it is in the fat rat, but not in early hyperphagia before obesity has developed. I have shown repeatedly that soon after operation, or after a period of restricted feeding, the operated rat will work harder for food than a normal rat. However there is another fallacy involved in the use of the totally unphysiological concept of hunger drive which the psychologists have invented. To measure drive, one must measure some sort of activity, and the assumption that hunger and activity have the same direct relationship in the hypothalamic animal as they have in the intact one is not valid. This is illustrated in the following experiments.

Brobeck showed, and we have confirmed, that hypothalamic obesity is a graded effect, the degree of which is a function more of the position in which a lesion is placed in the hypothalamus than the size of the lesion. Miss Bruce and I showed (Bruce and Kennedy, 1951) that a similar graded effect could be demonstrated on sex behaviour, but not on the endocrine control of sex cycle. The majority of fat rats showed no disturbance of oestrus rhythm or ovarian structure except in the late stages of obesity. But such rats would not mate. The interference with mating was absolute in the fattest rats, with lesions near the ventromedial nuclei, and only partial in mildly fat rats with lesions in other parts of the hypothalamus. I have shown, with Dr. Curt Richter, that an exactly similar relationship can be demonstrated between obesity and running activity as measured in the treadmill type of activity cage (unpublished data). All

our obese rats became inactive, but reduction of their weight by starvation restored activity in the mildly obese ones, while it was without effect on the fattest ones. The maximum effect both on food intake and the two different forms of activity was achieved by lesions in exactly the same place. It seems likely that this is one of the areas which facilitates activity in general. But it clearly suppresses feeding, which as Brobeck has suggested, is facilitated by the lateral hypothalamic area. So we have a reciprocal control at the hypothalamic level, of functions which are directly associated in the intact animal. This is not really surprising when one considers the work of Bard and Mountcastle (1948) on the role of this area of the hypothalamus in the related problem of sham rage and placidity. They have shown that still further facilitation and suppression areas in the cortex and amygdala influence the hypothalamic centres, and the whole control is probably a complex one.

Current theories of appetite seem to have become dominated by hypothalamic appetite. I have tried to show that the dominance of this type of appetite is characteristic only of the animal living on the edge of starvation, and becomes progressively less as we get bigger or fatter. Many different qualities of food must affect appetite. The reaction against the classical theories of Cannon and Carlson has gone too far. These authors were right to point out that as feeding is intermittent, a meal is over before changes in the level of blood sugar can stimulate central receptors, and that peripheral receptors which can be stimulated during a meal must play a part in satiety. The importance of psychogenic factors in obesity needs no restating, and is only one illustration of the importance of the cortical factors in appetite which make the gourmet. There is still a great deal to do before we can produce an integrated picture of the control of appetite in man, or an accurate assessment of the relative importance of appetite and activity in the aetiology of obesity. Meanwhile we may feel thankful to belong to a species which has to some extent emerged from the tyranny of the hypothalamus, even if our delight in food does sometimes involve the penalty of obesity.

References

Adolph, E. F. (1947) : *Amer. J. Physiol.* **151,** 116.
Anand, Bal K., and Brobeck, J. R. (1951) : *Yale J. Biol. Med.* **24,** 123.
Bard, P., and Mountcastle, V. B. (1948) : *Res. Publ. Ass. nerv. ment. Dis.* **13,** 107.
Bruce, H. M., and Kennedy, G. C. (1951) : *Proc. Soc. Study of Fertility III.*
Kennedy, G. C. (1950) : *Proc. roy. Soc.*, B, **137,** 535.
Kennedy, G. C. (1951) : *Proc. roy. Soc. Med.* **44,** 899.
Kennedy, G. C. (1953) : *Proc. roy. Soc.*, B, **140,** 578.
Miller, N. E., Bailey, C. J., and Stevenson, J. A. F. (1950) : *Science*, **112,** 256.

Deposition of Fat in Farm Animals

by

Dr. J. Hammond, C.B.E., F.R.S.

The composition of the body changes as the animal grows up. This change is brought about by the different tissues growing at different rates. Complete dissection data on pigs (McMeekan, 1940) and sheep (Pálsson and Vergés, 1952) have shown that the different tissues mature in the following order—nerve, bone, muscle and fat. Accentuation of any one of these tissues gives rise to the types described by Tanner in man.

As shown by various workers (McMeekan, 1940 ; Pomeroy, 1941 ; Pálsson and Vergés, 1952) by rearing animals on high and low planes of nutrition, the early maturing tissues have priority of supply of nutrients from the blood stream over the later maturing tissues.

Improvement of animals for early maturity, i.e. for putting on fat at an early age, has been brought about by rearing them on a high plane nutrition and then picking out and breeding from those which show a high fat to bone ratio at an early age. Thus the various breeds of sheep, for example, vary in the amount of fat they will put on at any one age. For example, the Southdown breed has not only a greater amount of fat than the Blackface Mountain in early life but it also has a larger proportion of fat to muscle in adult life.

The effect of low plane nutrition on an early maturing breed is to make it into a late maturing breed by delaying the time between the maximum growth rate of the different tissues.

The rate of growth, as measured in increase per unit of time, falls rapidly as

the animal ages. The rate of this fall varies in the different tissues, in nerve and bone it is earlier than in fat, and in muscle it is earlier than in fat. Clausen and his co-workers have shown (1954) in pigs that during the growing period there is a decrease in the daily production of muscle when a live weight of about 70 kg. is reached, whereas the maximum daily production of fat is not reached until about 100 kg. live weight and that this rate is maintained for some time. It should follow that if the animal is maintained on a low plane of nutrition until the rate of bone growth has sunk to a low level and then the animal is raised to a high plane of nutrition little further bone growth will be made, but large quantities of fat will be put on a small skeletal frame-work.

This is in fact what does happen, as experiments by McMeekan (1940) on pigs and Pálsson and Vergés (1952) on sheep have shown. Comparisons of animals from the same litter show that a pig fed on a high plane of nutrition while it is young and low plane later has a large skeletal frame-work and little fat while a pig fed on a low plane of nutrition when it is young and on a high plane later has a small skeletal frame-work with much fat on it.

If after such pigs have reached 200 lb. live weight they are both put on the same normal ration they will retain these characteristics throughout their life. In other words the pig with the large skeletal frame-work and early growing parts of the body (heart, liver, lungs, etc.) will have a higher maintenance requirement and so on the same ration will have less surplus nutrients to put on fat than will the small framed animal.

The most common cause of obesity in man is in our opinion similar to that which is described above for animals. Exceptionally only is it due to genetic mutations such as occur in mice.

References

Clausen, Johs Jesperson, R. Nörtoft Thomsen, and Winter, J. E.: 'The anatomic and chemical composition of pigs,' *Nat. Res. Inst. of Animal Husbandry*. Copenhagen.

Hammond, J. (1943): *Proc. Nutrition Soc.* **2**, 8.

Hammond, J. (1952): *Ann. Nutrition et l'Alimentation*, **6**, C. 119.

Hirzel, R. (1939): *Onderstepoort J. Vet. Sci.* **12**, 381.

McMeekan, C. P. (1940): *J. agric. Sci.* **30**, 276.

Pálsson, H., and Vergés, J. B. (1951): *J. agric. Sci.* **41**, 1.

Pomeroy, R. W. (1941): *J. agric. Sci.* **31**, 50.

EXPLORATION OF CAVES IN THE DERBYSHIRE DISTRICT

INTERIM REPORT No. 26 (1954) of COMMITTEE appointed to co-operate with a Committee of the Royal Anthropological Institute in the exploration of caves in the Derbyshire district. Mr. M. C. BURKITT, *Chairman*; Mr. A. LESLIE ARMSTRONG, *Secretary*; Dr. ARTHUR COURT, Prof. H. J. FLEURE, F.R.S., Mr. W. H. HANBURY, Dr. J. WILFRED JACKSON, Mr. R. U. SAYCE.

Since the submission of the 1953 Report the excavation of Ash Tree Cave, near Whitwell, has been continued by Mr. Leslie Armstrong, Dr. Arthur Court, Mr. W. H. Hanbury and Mr. Horace Jones and search has been made in the district for further sites of possible Creswellian occupation.

With regard to Ash Tree Cave, Mr. Leslie Armstrong reports that the excavation of the central section has been sunk to a depth of 12 feet over an area 11 feet 6 inches in width and 6 feet in breadth. Evidence of periodic occupation by Mousterian man persists and at the level now reached appears to be increasing. The cave earth has become much more calcareous, closely compacted, and full of small stones; but large slabs and rocks, which in the upper layers were so abundant, are infrequent. On both sides the cave wall is shelving rapidly backwards, thereby increasing the width of the excavation.

The artifacts recovered include a well-worked Mousterian point in flint, patinated white, similar in type to examples from the Pin Hole Cave, Cresswell; a few small flakes of flint; split quartzite pebbles; a number of large quartzite pebbles, possibly used as pounders, but bearing no evidence of use.

Small pebbles, about the size of acorns, have been numerous and may possibly be sling stones. They have been noted at all levels, but are of more frequent occurrence in this zone.

Bone tools and utilised split bones have been numerous and two large pieces of split bone, probably rhinoceros, or mammoth long bones, and a portion of rib, showing signs of use, have been found. The faunal remains chiefly comprise teeth; large bones are rare and fragmentary. Several teeth of woolly rhinoceros, both young and old, have occurred; also part of an uncut tooth of mammoth, teeth of hyaena, giant deer, horse, wolf and a number of cave bear teeth. Teeth of reindeer, or fragments of antler have been few, as have also rodent remains and molluscs.

Owing to Mr. Armstrong's absence in Rhodesia the work at Ash Tree Cave was discontinued during the summer, but his colleagues have made a valuable reconnaissance of the Whitwell and Whaley area, in search of possible rock shelters of Creswellian occupation and trial sections have been dug at three sites. Two of these, a rock shelter in a limestone outcrop, near the Parish tip at Elmton, and a site in proximity to Ash Tree Cave, yielded negative results, but a section on the line of the Whaley Cave and Rock Shelter (No. 2) outcrop is in progress and the indications are more promising there, but not yet positive.

During the progress of the Elm Tree Cave work facilities have been given this season to several students of archaeology to participate in the work of excavation and gain experience in the technique of cave digging.

The Committee desires to be re-appointed.

M. C. BURKITT,
Chairman.

A. L. ARMSTRONG,
Secretary.

August 14, 1954.

SECTION E.—GEOGRAPHY

DEMOGRAPHY IN INTER-TROPICAL AFRICA

FINAL REPORT of COMMITTEE appointed to collect and record information on demography and seasonal activities in relation to environment in Inter-tropical Africa. Mr. R. W. STEEL, *Secretary*; Prof. S. J. K. BAKER, Mr. K. BUCHANAN, Dr. R. J. HARRISON CHURCH, Prof. F. DEBENHAM, O.B.E., Prof. C. DARYLL FORDE, Prof. B. J. GARNIER, Dr. E. J. HOWELL, Mr. D. B. MATHER, Dr. R. MILLER, Prof. W. J. VARLEY, Prof. J. H. WELLINGTON.

The Committee records with deep regret the sudden death of its Chairman, Professor A. G. Ogilvie, in February 1954. Prof. Ogilvie was associated with the pre-war Committee from its establishment at the Oxford Meeting of the Association in 1926 and served as its Secretary until the war. But for his efforts, the Committee would not have been revived after the war, and its members are very conscious of the help and encouragement that he always gave it. It is a cause of much satisfaction that he was able to take the chair at the meeting of the Committee held at Keele in January 1954, only a few weeks before his death, and that it was then clear to him that, under his leadership, its work was almost complete.

At a meeting of the Committee of Section E held on September 8, 1953, approval was given to the Research Committee's proposal that it should produce (*a*) the bibliographical list of work on Tropical Africa published since 1945 by Commonwealth authors; (*b*) a list of Commonwealth geographers engaged in research on tropical African subjects, arranged regionally and with an indication of the subjects of research; and (*c*) some critical appreciation of the state of geographical research in tropical Africa at the present time. During the year members of the Research Committee have made further contributions to the two lists, and it is hoped that these will be completed before the Council of the British Association meets in November 1954, so that the Council may then consider the appropriateness of printing and circulating the report at the Association's expense. (The Section Committee made this request last year but the Council can consider such requests only when completed reports are before them. The Research Committee trusts that the Section Committee will be willing to repeat the request this year.)

Prof. Ogilvie had agreed to write the critical appreciation of the state of geographical research in tropical Africa, but in view of his death the Committee will have to meet and reconsider this section of the final report.

During the year the Secretary has corresponded with a number of students and others in Africa and in British universities and has been able to advise them on matters of geographical research. A note about the work of the Committee was published, through the kindness of the Editor, in a recent issue of *Geography*.

We have been in touch with the Twentieth Century Fund Survey of Africa, under the direction of Dr. George H. T. Kimble, through the International African Institute, and have decided that there is need for our proposed bibliographical list of specifically geographical material as well as for the more comprehensive bibliography that is being prepared by the Survey. An exchange of information has been arranged, and some members of the Committee are assisting Prof. Daryll Forde in the collection of matter for the Survey's bibliography.

The Committee regrets that, for various reasons, its work is not finished as it had been hoped that it might be a year ago, but as the bulk of the project has been completed it does not seek reappointment for a further year. It hopes that the Section Committee will commend its labours to the Council and that the Council will later see its way to producing and circulating its reports. The members of the Committee, both in Britain and in Africa, wish to thank the Recorders, Prof. D. L. Linton and Dr. A. E. Moodie for their encouragement and guidance and the members of the Section Committee who have shown interest and given support over a period of years. We hope that our report, when complete, will be a useful assessment of the work undertaken by geographers in tropical Africa during the post-war years.

R. W. STEEL, *Secretary*.

1954.

MEGALOSAURUS, THE OXFORD DINOSAUR [1]

by

DR. W. E. SWINTON

IT would be a piece of palaeontological negligence if on the occasion of the Association's visit to Oxford we were to forget the presence in these Museum buildings of the scientific foundation of the Dinosaurs. For although it was in the spring of 1822 that Mrs. Gideon Mantell found the first recorded and recognisable dinosaurian specimens, they were not scientifically named until 1825, and by that time Professor Buckland had already described the genus *Megalosaurus*. The basis for this name was a lower jaw, some vertebrae and parts of ribs, the pelvic bones, and the right femur, that had come from the Stonesfield Slate, at Stonesfield, about three miles west of Woodstock, and thus about nine miles N.N.W. of the Museum, where they came to be examined and where they remain.

Cuvier postulated 'a length exceeding forty feet and a bulk equal to that of an elephant seven feet high' for an individual of which the Museum has a femur measuring 2 feet 9 inches.

Comparable bones were also by now being found at Cuckfield in Sussex by Dr. Gideon Mantell himself and one of these 'nearly 20 inches in circumference' was thought by Buckland to be so massive as to warrant the conclusion that the original animal had been at least 60 to 70 feet long. On all this evidence, in which no doubt Mantell's specimen played some part, Buckland (in collaboration with his friend Conybeare) invented the name *Megalosaurus*. These original bones are freely accessible to the student in the University Museum in Oxford, but there appears to be no sign of the femur in the Mantellian collection purchased by the Trustees of the British Museum in 1839 and now housed in the Geological Department of that museum at South Kensington.

The bones of the Megalosaur were said to be accompanied in their original deposits by bones of long-legged birds allied to the Order Grallae 'thus affording the most ancient example yet discovered of the occurrence of birds.' [2]

The first view was that *Megalosaurus* was an amphibious animal, an idea supported by its being found with the remains of turtles, crocodiles and plesiosaurs.

Through the course of the years numerous other bones of *Megalosaurus*, attributed to the species *M. bucklandi* established by Hermann von Meyer in 1832, have been discovered, but the very plenitude of material has been a difficulty since so little of the skeleton has been found associated and most of the specimens have been very fragmentary.

In 1913, the collection of Mr. James Parker came to the Oxford University Museum, a series of bones of great interest, and containing many Megalosaurian specimens.

All this material has been examined by Freiherr von Huene in the comprehensive study of the family Megalosauridae that has been undertaken and based on the evidence of 96 separate finds of material in Europe and of 15 extra-European alleged occurrences. As a result of this survey this distinguished palaeontologist recognised four species: *Megalosaurus bucklandi* von Meyer; *Streptospondylus cuvieri* Owen; *Poikilopleuron bucklandi* Deslongchamps; and *Megalosaurus bradleyi* Smith Woodward. Of these the typical Jurassic carnosaur is represented by *M. bucklandi*.

The general characters of this form may be summarised as: Large carnivorous dinosaur, the femur being 1 metre long; premaxilla high and short; maxilla long

[1] Paper delivered to Section C (Geology) on September 7, 1954, at the Oxford Meeting of the British Association.

[2] These were Pterodactyl bones.

and narrow, divided almost into halves by the ascending (nasal) portion, the latter being bent posteriorly at 90 degrees; hinder alveolar border of maxilla concave; narial cavity large; only one preorbital opening, probably half the size of the orbit. Alveolar margin of lower jaw almost straight. Larger teeth at anterior end of jaws; teeth serrated, on hinder edge and on upper half of anterior edge; four premaxillary teeth followed by ? 14 maxillary teeth.

Cervical vertebrae opisthocoelous. Dorsals with pleurocentral cavities not so deep as in *Streptospondylus* or *Allosaurus*. Length of dorsal neural spines above diapophyses approximately equal to length of centrum.

Scapula long, straight and narrow; average breadth about one-sixth the length. Scapula plus coracoid as long as femur. Pubis rod-like. Distal end of femur much curved. Hind foot functionally three-toed, with first toe reduced and fifth toe probably absent. Astragalus with ascending process.

The posterior parts of the skull are largely unknown, but it would appear that the quadrate was directed downwards and backwards so that the gape of the jaws was wide. This, with the sharp, recurved and laterally compressed teeth, would make the animal a formidable carnivore. Coupled with its large size and the undoubted bipedal ability of the strong hind legs, *Megalosaurus* presented an impressive appearance and the size of the animal, if less than that attributed to it by the first collectors, must often have exceeded 20 feet as measured from the snout to the end of the tail. Though the hands and feet had digits with strong and curved claws, the forelimbs were much smaller than the hind and could not have been used in progression, but they would be used in grasping and tearing the prey.

The pose of this considerable and heavy animal must have been less erect than is usually depicted on the restorations, with the pubis consequently approaching the angle of 35 degrees to the vertical in direction rather than the 70 degrees more usually shown. It is significant that the distal end of the pubis shows a slight enlargement, foreshadowing the great rocker-like process characteristic of the later and larger forms.

The first restoration of a Megalosaur was that made by Waterhouse Hawkins under the supervision of Richard Owen, showing the animal on all fours, and made for the Geological Island in the Crystal Palace Grounds in 1853.

By that date another species with Megalosaurian affinities had been described, the so-called *Streptospondylus cuvieri*. The history of this species is worth recalling. A series of bones, none giving a satisfactory picture of the animal to which they belonged, had been discovered in the Upper Callovian of Honfleur, near Trouville in Normandy. They had been described, but not named, by Cuvier, who none the less was impressed by their crocodilian affinities. The generic name *Streptospondylus* was given to them by von Meyer in 1832. When Richard Owen came to write his great paper on the History of Fossil Reptiles in 1841 he gave a brief résumé of the characters of this genus and proceeded to define the species, *S. cuvieri*. His types for the purpose were half an anterior dorsal vertebral centrum, then in the possession of a Mr. Kingdon, and said to have come from the Inferior Oolite of Chipping Norton; a smoothly worn and incomplete tooth; and a vertebra from the Lias (presumably Upper Lias) of Whitby, then in Mr. Ripley's collection. All of these three specimens are no longer to be found. The tooth is most easily dismissed for it seems to have had all the features of a worn Megalosaurian tooth. There seems little likely relationship between the incomplete vertebra from the Inferior Oolite and the rather different vertebra from the Upper Lias. Indeed, since Owen realised the affinity of the specimens with *Megalosaurus*, yet placed them in a new species of an allegedly crocodilian genus, it might at first sight be thought that the basis of the species is lacking in determinants. This view is also borne out by Phillips, who states in his *Geology of Oxford* (p. 322), that he thinks the Chipping Norton vertebra to be that of a young *Ceteosaurus*. But it is not so. Owen gives a most detailed description of the structure of the

vertebral diapophysial structure and support and the shape and apparent characters of the neural spine. More important, he is precise in the location and dimensions of the deep pit behind each costal surface, which appears ' as if a man had pressed his two thumbs forwards and inwards up to the first joint into the substance of the body of the vertebra ' (1841, p. 88). Such characters in opisthocoelous vertebrae of the cervical and anterior dorsal regions strongly recall, as, indeed, they did to Owen, the condition of the Honfleur specimens, which are still available for study in the Jardin des Plantes in Paris.

There the matter of this species rested until, in 1913, there came to the University Museum in Oxford a fine series of bones of a small dinosaur, and apparently all from the same individual, in the collection of Mr. James Parker. His specimens had been obtained from the Oxford Clay of Summertown in North Oxford (and not on the ' South Coast of England,' as stated by von Huene in 1926). It was not possible to compare the new materials with the types of Owen's description, for they were already lost, but the Summertown specimens agree very closely with the Honfleur ones, which Owen had regarded as very close to his types in 1841 and one specimen of which was figured in his *Palaeontology* (1861, p. 299, fig. 103).

The specimens in the Parker collection allow a reconstruction of a great deal of the skeleton to be made and, indeed, constitute the most complete Jurassic carnosaur in Europe. This skeleton, that of a small and certainly young Megalosaur, had already been studied in some detail by Baron F. Nopcsa (1905), and in much greater detail by Freiherr von Huene (1926). It is now possible, therefore, to give a brief account of the characters of this species for comparison with those already given for *M. bucklandi*.

M. cuvieri (Owen) syn. *Streptospondylus cuvieri* Owen. Skull with one antorbital fenestra. (Nopcsa states that two were probably present but there is no evidence of this.) Ascending process of maxilla very narrow ; alveolar border of maxilla flat ; alveolar border of mandible widely concave. Four teeth on premaxilla ; 12 teeth in maxilla ; 14 alveoli in lower jaw. All the teeth are shorter and broader than those of *M. bucklandi*. The vertebral series is well represented by probably the 2nd–9th cervicals ; 1st–14th dorsals ; 3 undoubted sacrals ; and 29 caudals. The cervicals and anterior dorsals have the general characters of the Honfleur specimens and the pleurocentral cavities are marked.

In the shoulder girdle the right scapula is entire or nearly so and shows the same length and straightness that marked that of *M. bucklandi*, though the breadth is contained in the length seven times, and the bone is thus proportionately longer than that of the former species. Nopcsa considers the two scapulae to be similar.

Bearing in mind the youth of this specimen of *M. cuvieri*, the humeri of the two species are not markedly different, though the much larger radial crest in *M. cuvieri* is noteworthy.

The pelvic elements show a variety of affinities that serve to consolidate the specific relationship of the Oxford and Honfleur specimens. The right ilium of *M. cuvieri*, for example, is very similar, apart from its size, to that of *M. bucklandi* ; and the ischium is almost identical with that of the Honfleur specimen. The pubis is complete, is less distally expanded than Nopcsa asserts, and, in the present writer's opinion, is wrongly mounted, being much too elevated and anteriorly directed. This is largely because in the mounted specimen the pubic articulation with the ilium is artificial. None the less, the pubis is 41 cm. long and thus has an almost identical ratio to the length of the femur as these bones have in *M. bucklandi*. The femur is straighter than that of *M. bucklandi* and the distal condyles are more widely spaced. In this young specimen the femur exceeds the tibia by only 2 cm., being 52 cm. in length. The astragalus has an ascending process, similar to that in the Honfleur specimens, but not proportionately so high as that of *M. bucklandi*. The sacrum probably had 5 vertebrae. The hind foot was four-toed.

Restorations of the skeleton have been made by Nopcsa and by von Huene, the former showing less material than actually

exists and showing the animal with stronger forelimbs and a more generally upright character than von Huene depicts. Of the two, recent examination would seem to support von Huene's restoration and, in fact, to go further in suggesting that even that has too upright a pose. The weight distribution, the comparative shortness and weakness of the forelimb, coupled with the development of the long counter-balance of the tail, all suggest to the writer that *Megalosaurus* had a much more stooping gait.

The relationship between the young *M. cuvieri* and the adult *M. bucklandi* may be questioned. The answer is that the characters of the young specimen are borne out with little significant alteration in the Honflueur specimens, which are certainly in some cases half as large again. The characters of the forelimb ; the differently shaped and comparatively longer femur of *M. cuiveri*, and the slightly different characters of the pelvis all tend to establish the specific identity of the last-named species.

Nopcsa (1906, p. 293), suggests that the smaller species may have attained a total length of 20 feet and 27 feet for the Oxford and the Honfleur specimens respectively, although the former as now mounted in the University Museum is slightly less. Nopcsa estimated the length of *M. bucklandi* as attaining ' a maximum length of 36 feet,' which is probably somewhat exaggerated.

Several other British species of *Megalosaurus* have been described but few of these can be strictly contained within the genus to-day. Several bones, including three dorsal vertebrae with high neural spines, also in Mr. Parker's collection in the Oxford Museum, but which originally came from the Oxford Clay just north of Weymouth, have been made the types of a new species, *M. parkeri* by von Huene (1926, p. 76). They already show a heightening of the spines that reached a maximum for this group in *M. dunkeri* Lydekker, and this species has been transferred to a new genus *Altispinax* by von Huene (1926, p. 78). The older specific names, *M. oweni* Lydekker, 1889, and *M. insignis* Eudes Deslongchamps, 1870, are too little supported by evidence for any close determination of their real relationships, though some little known bones in the Parker collection, from the Inferior Oolite at Nethercombe, near Dorset, have been made the types of *M. nethercombensis* von Huene (1926, p. 72), which he later (1932, p. 220), transferred to a genus *Magnosaurus* (1932).

One of the most outstanding carnosaur discoveries in this country was the skull from the Great Oolite of Minchinhampton, S.S.E. of Stroud, in Gloucestershire, named *Megalosaurus bradleyi* by Smith Woodward (1910, p. 110). This specimen has a well-developed nasal horn. Since the skull is only 26 cm. long and has not the general shape or characters of *M. bucklandi* or *M. cuvieri*, but, on the contrary, seems to have close relationship with the American carnosaur *Ceratosaurus*, there are good grounds for accepting it as the type of a new genus, *Proceratosaurus* von Huene (1926, p. 70).

The problems aroused by these tantalising and incomplete remains and the equal difficulties sometimes caused by nomenclatural insufficiency are obvious enough, but since von Huene's revision of the group many other occurrences of *Megalosaurus* have been alleged, and I have refrained from mentioning those species of more ancient time whose types are on the continent of Europe.

In recent years, A. F. de Lapparent (1943, 1947) has recognised *Megalosaurus insignis* Eudes Deslongchamps in the Jurassic of Damparis (Jura) and *M. pannonniensis* Seeley in the Danian of the south of France. Already the genus had been recognised in the Lower Cretaceous of Tendaguru in Tanganyika (*M. ingens* Janensch) ; while *M. crenatissimus* Déperet, *M. saharicus* Depéret, and *M. superbus* Sauvage had all been re-identified or named from the Cretaceous. The field has been extended by Rana Rao's suggestion (1932), that *Megalosaurus* occurs in the Upper Cretaceous of South India, though the tooth and the ? cervical vertebra he describes do not wholly substantiate the claim. Much more serious consideration must be given to the discovery by S. Welles of an almost complete skeleton from the Middle Jurassic of Arizona, and which he has named *Megalosaurus wetherilli* (1954). Welles argues a close relationship largely as a result of

statistical analysis of bone proportions but the excellent illustration of the animal shows that the main elements of the skull, the scapula, and the pelvic bones, are not those of typical Megalosaurs. This new skeleton has, however, somewhat similar proportions of its limbs and a well developed processus ascendens in the astragalus. The characters of the cervicals are not visible and are not described. Its relationships are certainly with the early carnosaurs, but it is not *Megalosaurus*.

Re-examination of all the material that is now labelled as *Megalosaurus*, is highly desirable and in such a study it is clear that the type material in the University Museum, Oxford, will continue to provide the standard for comparison and nomenclatural certainty.

References

Buckland, W. (1824): *Trans. Geol. Soc. Lond.*, Ser. 2, **1**, 390.

Deslongchamps, E. E. (1870): In Lennier, *Etudes Geol. et Pal. Embouchure Seine*, p. 35.

Cuvier, G. (1834): *Ann. du. Mus.* **12**, 73.

Huene, F. F. von (1926): *Revista del Museo de la Plata*, **29**, 35–167.

Huene, F. F. von (1932): *Monog. Geol. Palaeont.*, Pt. I, p. 220.

Lapparent, A. F. de (1943): *Mem. Soc. géol. France*, N.S. XXI, Mem. No. 47.

Lapparent, A. D. de (1947): *Ibid.*, N.S. XXVI, Mem. No. 56.

Lydekker, R. (1889): *Geol. Mag.*, Ser. 3, **6**, 325.

Meyer, H. von (1832): *Palaeologica*, p. 106.

Nopcsa, Baron F., (1905): *Geol. Mag.*, N. Ser. Decade V, **2**, 289, July, 1905.

Owen, R. (1842): *Rep. Brit. Assoc.* (1841), p. 88.

Owen, R. (1861): *Palaeontology*, Edinburgh.

Phillips, J. (1871): *Geology of Oxford*, Oxford.

Rao, L. Rama (1932): *Amer. Journ. Sci.*, **24**, 221.

Welles, S. (1954): *Bull. Geol. Soc. Amer.*, **65**, 591.